Design of ferrocene-peptide dimers for DNA binding and electrochemical sensing

by

Xun Zhong

UNIVERSITY OF BIRMINGHAM

A thesis submitted to
The University of Birmingham
For the degree of
DOCTOR OF PHILOSOPHY

School of Chemistry
College of Engineering and Physical Sciences
University of Birmingham
September 2013

Abstract

GCN4 is a yeast transcriptional factor, which binds sequence selectively to DNA as a homodimer. Two domains are required to achieve this, an N-terminal basic domain, which interacts with the target site directly, and the C-terminal leucine zipper responsible for dimerisation. The latter can be replaced by chemically linking together two basic domains, to retain similar DNA binding. This work investigates using a ferrocene complex as the dimerization unit due to its flexibility and its characteristic electrochemical redox properties.

A strategy was proposed which took advantage of coupling to the thiol side chains of two cysteine residues, and detailed studies on dimerising two equivalents of Cys, or the tripeptide glutathione, were performed.

Design of five peptides based on GCN4, which were synthesized and dimerized with a ferrocene linker molecule, to combine the attractive redox electrochemistry of the stable ferrocene and the DNA binding ability of GCN4 dimerized basic domain. Molecular dynamic models were produced for four of the ferrocene peptide dimer complexes bound to target DNA over 10 ns. Their DNA binding ability was investigated by circular dichroism, and electrochemistry was studied in the absence and presence of non-specific and target DNA respectively.

Table of contents

Acknowledgements

I would like to thank my supervisor Dr Anna Peacock for her continuous support over the past four years, without her guidance this project would not have been possible. I would like to give thanks to the analytical facility for advice on NMR, MS and chromatography advices. I am very grateful to Emmanuel Oheix, Matt Berwick, Gemma Bullen and Dr Aimee Gamble for their help and advice in my experimental studies; Professor James Tucker's group for the preparation of DNA; Dr Sarah Horswell and Goy for their help with electrochemistry; and Dr John Wilkie for his help in MD simulations. Finally would like to thank my family for their financial and mental support, without them I would not be here today.

List of Abbreviations

DNA	deoxyribonucleic acid
Cys	cysteine
Tyr	tyrosine
Phe	phenylalanine
Tyr	tyrosine
Leu	leucine
His	histidine
Asn	asparagine
Ala	alanine
Lys	lysine
Gly	glycine
Asp	aspartic acid
Val	valine
Glu	glutamic acid
Ser	serine
SCE	saturated calomel electrode
UV	ultraviolet

UV-vis	ultraviolet and visible
PSA	prostate specific antigen
PR	protease
PNA	peptide nucleic acid
ITO	indium tin oxide electrode
PAH	poly (allylamine hydrochloride)
CD	circular dichroism
SPPS	solid phase peptide synthesis
CV	cyclic voltammetry
DPV	differential pulse voltammetry
Fmoc	fluorenylmethyloxycarbonyl
TFA	trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
HBTU	2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium
DIPEA	N,N-diisopropylethylamine
EDT	1,2-ethanedithiol
Pbf	pentamethyldihydrobenzofuran
A	absorbance

ε	molar absorptivity
c	solute molar concentration
l	sample cell length in cm
NTB^-	2-nitro-5-thiobenzoate
NMR	nuclear magnetic resonance
MS	mass spectrometry
EI	electron impact
ESI-MS	electrospray ionization mass spectrometry
MALDI-MS	matrix-assisted laser desorption ionization mass spectrometry
HPLC	reversed phase high-performance liquid chromatography
θ_{MRE}	molar residual ellipticity
c	concentration
d	cell path length in cm
n	number of residues in the peptide
E_{intial}	initial potential
$E_{switching}$	switching potential
i_{pa}	anodic peak current
E_{pa}	anodic peak potential

i_{pc}	cathodic peak current
E_{pc}	cathodic peak potential
k_{ET}	rate constant of the electron transfer process
k^o	rate constant for standard heterogeneous electron transfer
α	transfer coefficient
n	number of electrons transferred for each molecule
R	universal gas constant
T	temperature in K
η	overpotential
$E^{O'}$	formal potential
$E_{1/2}$	redox potential
F	Faraday constant
i	peak current (A)
A	surface area of the electrode (cm)
D	diffusion coefficient of diffusion species (cm^2/s)
C_o	concentration (mol/cm^3)
t	time in second
DCM	dichloromethane

GSH	glutathione
diGS	di glutathione
diCys	di cysteine
AA	amino acid
HOBT	hydroxybenzotriazole
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
fc	ferrocene
Δ*E*p	difference between oxidation potential and reduction potential
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
MeCN	acetonitrile
MeOH	methanol
Ac	acetyl
Cp	cyclopentadienyl
DHP	9,10-dihydrophenanthrene- 9,10-diol

Chapter 1 Introduction and background

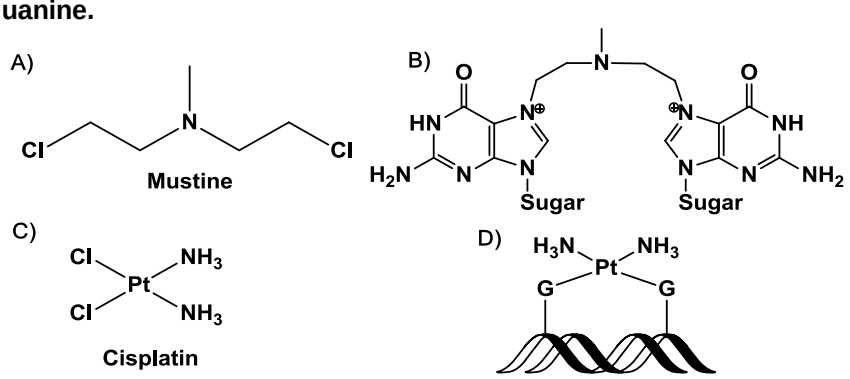
1.1 DNA binding molecules

DNA contains the genetic information of life and can adopt various three dimensional structures, the most common of which is B-DNA. In B-DNA, two strands are held together tightly as a double helix, which is stabilised by intermolecular hydrogen bonds formed between complementary base pairs as well as the hydrophobic interactions holding the apolar base surfaces away from the aqueous solution.[1, 2]

Binding to B-DNA provides a method for the inhibition of transcription. This can therefore ultimately be used to control gene expression. Binding to B-DNA can be achieved through alkylation or metalation by covalent or dative bond formation, or through intercalation or groove binding by the formation of non-covalent bonds.

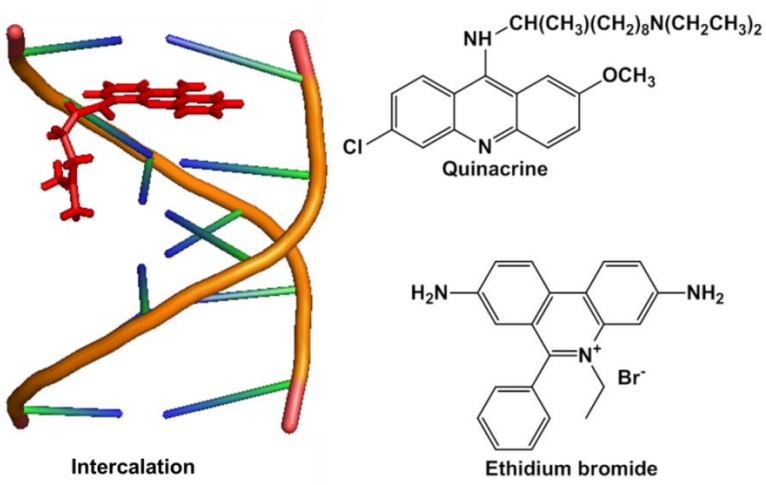
Agents which bind directly to DNA or alkylate it, have found applications in cancer chemotherapy.[3] DNA alkylating agents prefer to bind to N-7 of guanine but can also bind to N-3 of adenine and some alkylating agents such as nitrogen mustard (e.g. mustine, see Figure 1A), have two DNA binding sites and form cross bridges (see Figure 1B). Cis-platin is a widely used anti-cancer drug which binds to DNA through the formation of dative bonds, as it cross links the nitrogens of DNA base, see Figure 1C-1D.[4]

Figure 1 A) Mustine, B) the cross link formed by nitrogen mustard between the two guanine sites of DNA and this can be inter-strand, C) Cis-platin, D) the inter or intra-strand cross link by cis-platin, G:guanine.



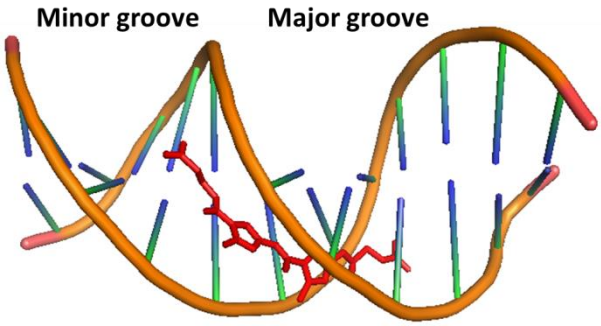
DNA intercalators bind to DNA by inserting planar aromatic rings between the DNA nucleobases, see Figure 2. The binding of intercalators to DNA is not sequence specific and is stabilised by π-π stacking, van der Waals and hydrophobic forces. Through not sequence specific, intercalation happens more in G/C rich sequences as they are more readily unstacked.[5]

Figure 2 Structure of two DNA intercalators: quinacrine and ethidium bromide, X-ray crystal structure of the acridine-DNA complex, acridine is shown in red (PDB code: 452D).[6]



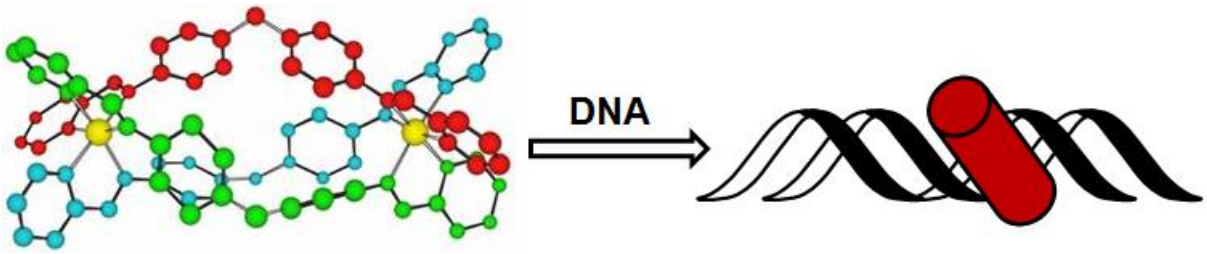
Grooves exist between the two strands and because these are not positioned directly opposite each other, the grooves are unequal in size, see Figure 3. The major groove is 22 Å wide, whereas the minor groove, is 12 Å wide.[7] These grooves are adjacent to the base pairs and provide potential binding sites. For example pentamidine and netropsin are minor groove DNA binding ligands, see Figure 3.[8,9]

Figure 3 X-ray crystal structure of the complex formed between DNA and the minor groove binder netropsin (shown in red) (PDB code: 101D).[9]



Hannon and co-workers prepared a tetracationic metallo-supramolecular cylinder $[Fe_2L_3]^{4+}$ which is a triple helicate through interaction of iron (II) salts with pyridylimine ligands, see Figure 4.[10] The cylinder has similar size with the recognition domain of zinc finger protein. The cylinder binds to the DNA major groove and induces intramolecular DNA coiling.

Figure 4 The $[Fe_2L_3]^{4+}$ cylinder binds to the DNA major groove, each ligand strand (L) is coloured to show its triple helical structure.



1.2 DNA binding proteins

Sequence specific DNA binding proteins generally bind to the major groove of B-DNA through an α-helix, as this has the correct shape and size to fit into this groove. Sequence specific binding is achieved due to the formation of favorable hydrophobic and electrostatic interactions, as well as specific hydrogen bonds between amino acid side chains and DNA nucleobases exposed in the major groove. The major DNA-binding protein motifs, the zinc finger, helix-turn-helix and leucine zipper will be described in the following sections.

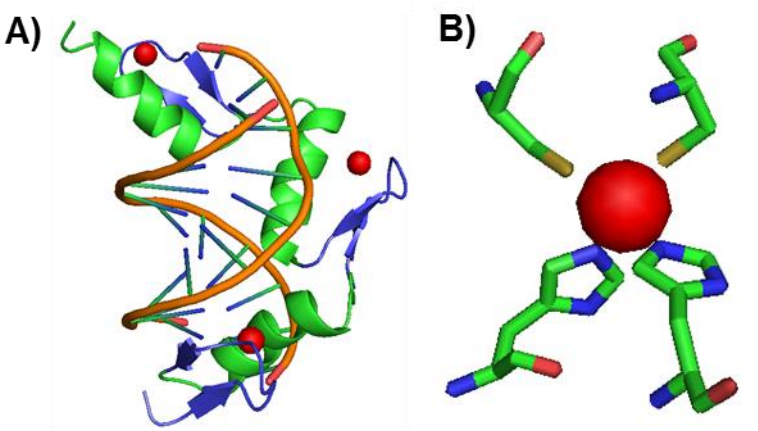
1.2.1 Zinc finger

The zinc finger, is a protein structural motif stabilised by the coordination of one or more zinc ions. There are different types of zinc finger proteins including Cys_2His_2 zinc finger, gag knuckle, treble clef, zinc ribbon, and Zn_2/Cys_6 zinc finger. However,

the Cys_2His_2 zinc finger, treble clef and zinc ribbons are the most common.[11] The family of the Cys_2His_2 zinc finger motif is the best-characterised class of zinc finger family, and contains many proteins with important functions. They are the most common DNA-binding motif encoded for in the human genome, as well as in some other multi-cellular organisms. The human genome contains about 700 Cys_2His_2 zinc finger coding genes. Though the majority binds to DNA, some zinc fingers also have RNA binding properties.[12,13]

The first zinc finger protein discovered was TFIIIA, a transcription factor with a repeat motif of 30 amino acids and a zinc binding site.[4] The repeating motif contains a conserved sequence of (Tyr Phe)–X–Cys–$X_{2,4}$–Cys–X_3–(Tyr, Phe)–X_5 –Leu–X_2–His–$X_{3–5}$–His (X represents any amino acid), in which Cys and His residues, which coordinate the zinc ion were very highly conserved.[5] NMR studies revealed that a Cys_2His_2 zinc finger consists of an internal αββ fold, which is stabilised by the chelatation of a zinc ion, and various interactions, including hydrogen bonding, electrostatic interactions and hydrophobic forces, between the conserved residues.[14,15] The zinc ion is coordinated by the conserved Cys in a β-turn and the conserved His from an α-helix.[14] The crystal structure of the DNA–Zif268 complex solved at 2.1 Å resolution, revealed that the amino acids at positions 1, 3 and 6 with respect to the start of the α-helix in the Cys_2His_2 zinc finger, contact DNA to recognise three continuous base pairs on a sequence of the target DNA. Additionally, later studies on the DNA–Zif268 complex solved at 1.6 Å resolution proved the importance of the residue at α-helix position 2 in DNA binding, due to the formation of a cross-strand interaction with a base on another DNA strand, see Figure 5.[16,17]

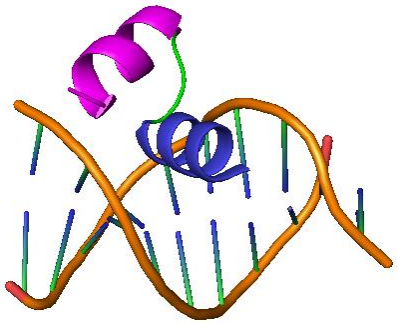
Figure 5 A) X-ray crystal structure of Cys_2His_2 zinc finger-DNA complex, the zinc is shown as a red sphere, the α-helix as a green ribbon and the β-turns in blue; B) an upclose of the Zn Cys_2His_2 coordination sphere (PDB code: 1ZAA).[16]



1.2.2 Helix-turn-Helix

The DNA binding helix-turn-helix structural motif consists of an α-helix, a turn region which is a short sequence of amino acids which links it with a second α-helix.[18,19] This motif is found in many proteins that regulate gene expression. In most cases, the second helix towards the C-terminal of the motif is involved in sequence-specific DNA base interactions, and so it is termed the recognition helix.[20] The recognition helix binds to the DNA major groove through hydrogen bonding, hydrophobic and electrostatic interactions with the exposed bases, see Figure 6.[21] The first α-helix stabilises the protein-DNA interaction, but does not play a direct role in the recognition.

Figure 6 The X-ray crystal structure of the bacteriophage lambda-Cro protein helix-turn-helix domain bond to DNA target site. The recognition helix located in the major groove of DNA is shown in blue, and the second α-helix in magentas (PDB code: 6CRO).[22]



1.2.3 Basic region leucine zipper

The basic region leucine zipper family of DNA binding proteins (bZip) was first identified by McKnight and co-workers in 1988.[23] They were found to bind specifically to target double stranded DNA in the major groove as a dimer. The basic domain of the bZip consists of an N-terminal 20-30 amino acid basic region which forms direct contacts with the DNA by binding to the major groove as an α-helix and forming multiple hydrogen bonds, electrostatic interactions and van der Waals forces with the exposed nucleobases (Figure 7). The C-terminal leucine zipper domain is characterised by a pattern of leucine residues repeating every 1st and 4th amino acid of a heptad repeat. The basic and leucine zipper domains are separated by a hinge area of 6 amino acids.[24,25] The leucine zipper region mediates dimerisation, by forming a two stranded and parallel coiled coil, whilst not interacting directly with the DNA, see Figure 6.[26-28] The bZip motif forms the focus of this project and will be introduced in detail in the following section, using GCN4 as an example.

Figure 7 X-ray crystal structure of the bZip domain of GCN4 bound to the CRE DNA target site, as homo dimer (pdb code: 2DGC), the leucine zipper domain is shown in yellow, the hinge domain is shown in green, basic domain is shown in red and the N terminal cap is shown in black.[29]

N-MIVPESSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER-C

220 Basic domain hinge Leucine zipper 281

1.3 GCN4

GCN4 is a basic leucine zipper eukaryotic transcriptional activator protein, which binds selectively to the promoters of yeast amino acid biosynthetic genes (Figure 7).[30] The GCN4 leucine zipper has been studied extensively with a large number of mutagenesis studies in the literature.[31-34] In addition, various peptides have been designed and synthesised based on the GCN4 leucine zipper motif, for DNA binding activity studies.[33-38]

GCN4 binds selectively to the 9 base pair CRE region, 5'-ATGACTCAT-3', of the HIS3 gene, which encodes for a histidine biosynthetic enzyme regulated in response to histidine starvation.[1,39,40] A region between 86 and 99 base pairs of HIS3 gene is important in positive regulation of HIS3 expression, and this region containing the TGACTC sequence is the HIS3 promoter region. GCN4 binds to the HIS3 regulatory sequence to activate their transcription.[41] GCN4 also binds to activator protein 1 (AP-

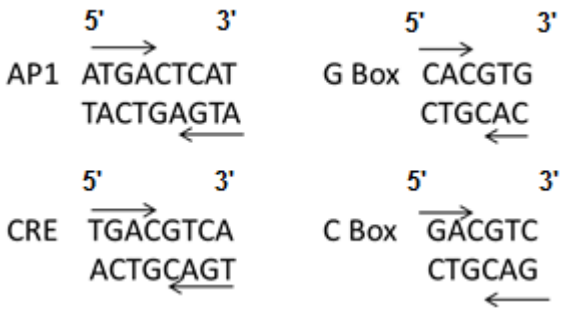
1), which is an important transcriptional regulator in disease pathways including the inflammatory response.[40]

Water molecules are an integral component of the DNA peptide interface in the complex formed between bZIP and DNA, as they mediate some of the interactions between the basic domain and DNA, and thus they contribute to stabilisation of the selective binding.[29]

Both the basic and leucine zipper domain of GCN4 are critical for achieving selective DNA binding, as is the case for the entire family of bZIP proteins.[42] When GCN4 binds to the target DNA, the DNA binding of GCN4 is symmetrically arranged and thus the activation of transcription on either side is equal.[33]

The GCN4 homodimer can bind to both the AP1 and CRE DNA site, which have symmetrical sequences and share the same half site but differ in their spacing.[43] Solution measurements indicate that DNA containing AP-1 is in a linear B form, whereas CRE is bent towards the major groove by 8° when bound to GCN4.[44,45] Many variations of these sequences can be efficiently bound by bZIP proteins.[43] For example,TGA1a have selective binding with C Box; EmBP-a, HBP1a, CPRF1, TAF1 and OCSBF1 have selective binding with G Box; CPRF2, TGA1b and GBF 1,2,3 have both selective binding with G Box and C Box. Interestingly, all these binding sites contain symmetrical sequences, see Figure 8.[43]

Figure 8 AP1, G-BOX, CRE and CBOX binding sites for bZIP proteins, all of which have symmetrical sequences. [43]

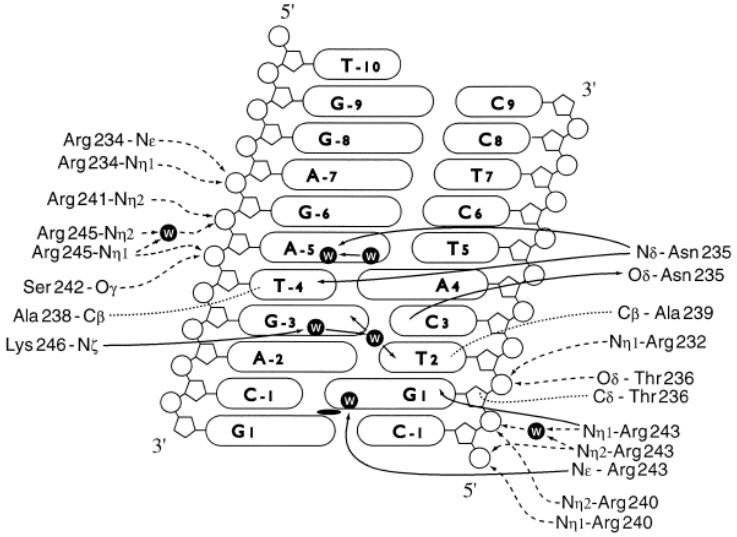


In the binding of GCN4 with the AP1 site, the GCN4 first contacts the central G in the AP1 site with the arginine residue in one monomer asymmetrically; this is followed by another four amino acids in each monomer binding to each half site symmetrically, in which asparagine (N235) at the center of the peptide forms hydrogen bonds with C and G in DNA, while alanine (A238 and A239) and serine (S242) also form hydrogen bonds with two T. GCN4 also binds selectively to the CRE site which contains an additional central G-C base pair, but with reduced binding affinity compared to the AP1 site. The conserved Arg in each GCN4 monomer most likely forms contacts with each G base in the first instance.[46] CRE can be considered as a modified AP-1 site, the addition central G-C base pair rotates the half sites by 36° and separates them by 3.4 Å.[47] Therefore the ability of GCN4 to bind with both AP-1 and CRE, requires the flexibility of the hinge domain between the leucine zipper and basic region, which has been proved by mutagenesis studies.[46-48] The binding preference for AP-1 and CRE sites is because of the residues in the hinge domain, which control the basic region helices orientation for alignment with either the AP-1 or the CRE site.[47] The mutagenesis studies indicate that residue Leu247, located in the hinge domain, is involved in this binding preference.[48]

Richmond and co-workers obtained the crystal structure of the GCN4 bZip domain-AP1 complex at 2.2 Å resolution.[29] In the basic domain of GCN4, five residues form direct contact with DNA bases in the major groove, while another seven residues form hydrogen bonds with the phosphate backbone. An important solvent network of 46 water molecules is involved in both types of interaction (Figure 9).

Figure 9 GCN4-CRE DNA interactions; the N-terminal basic domain residues form direct or water mediated interactions with the DNA bases and the phosphate backbone. The direct and water mediated interactions with nucleobases are shown as in solid lines; the interactions with the DNA phosphate backbone are shown as dash lines and hydrophobic interactions are shown as dotted lines. The directions of hydrogen bonds are shown in arrows. Figure adapted from reference 29, copyright 1995, with permission from Elsevier.

	230	235	243	250		
N-terminus	A<u>L</u>	KRAR<u>N</u>TEAARRS<u>R</u>ARK		LQR<u>M</u>KQ	Leucine zipper	C terminus



1.3.1 Hinge domain

Struhl and co-workers designed and synthesised peptides based on GCN4 with various amino acids inserted between the leucine zipper region and the hinge domain to study how these altered DNA binding.[49] Peptides with 2, 4, 6, and 7 amino acids inserted were prepared respectively. The 2, 4 and 6 amino acid insertion peptides were unable to bind to DNA, due to the spatial disruption of the protein, in which the basic domain extension from the leucine zipper is not maintained but instead is rotated. In contrast, most of the 7 residue insertion peptides bind with wild type affinity or slightly reduced affinity. If the peptide N-terminus continuously extends from the leucine zipper region, which is responsible for dimerisation, to the basic domain, the spatial relationship between the basic domain and the leucine zipper

subdomain is retained, as 7 amino acids correspond to a full heptad repeat and two full turns of an α-helix (3.6 residues per turn). The functional 7 residue insertion peptides form homo dimers. Though the 2 and 6 residue insertion peptides can dimerise, they cannot bind to DNA, due to the wrong spatial alignment of the basic domain in these mutants relative to wild type GCN4. The 7 residue insertion peptides can form DNA binding homo dimers, but can also dimerise with wild type GCN4 to yield hetero dimers. However, these hetero dimers cannot bind to DNA, as the spacing between basic region and leucine zipper of these peptides are not mutually compatible.[49]

1.3.2 Increasing α-helicity of the basic domain

The basic domain of GCN4 is not helical prior to binding to the DNA target site, however it becomes α-helical on binding.[33] Shin and co-workers simplified the basic region of GCN4 by replacing many of the amino acids with alanine. This results in a more α-helical structure and therefore allowed the alanine modified GCN4 to probe the relationship between helical content and DNA binding affinity, and to identify key amino acids crucial for sequence selective DNA binding. The sequences 4A, 11A and 18A have four, eleven, and eighteen alanine mutations in their basic region respectively, see Table 1.[50,51] The 11A and 18A sequences are also mutated in the hinge region, which separates the basic regions and leucine zipper regions of the monomers, and has an effect on DNA-binding.[52,53]

Table 1 Sequences of the basic region of GCN4, 4A, 11A and 18A, investigated by Shin and co-workers.[50] Alanine substitutions are underlined.

	Basic region	Hinge	Leucine zipper
	226	252	
Wild-type	DPAALKRARNTEAARRSRARK	LQRMKQ	
4A	ARAAAARARNTAAARRSRARK	LQRMKQ	
11A	ARAAAARARNTAAARRSRAAK	AAAAAA	
18A	AAAAAAAAANAAAAAAARAAK	AAAAAA	

The residues Asn[235], Ala[238], Ala[239], Arg[243] and Lys[246] were not substituted in the three peptides, because they are highly conserved among basic region leucine zippers, and can be involved in direct contact with the DNA bases in the crystal structures of the GCN4-DNA complexes. Furthermore Lys[246] is important for retaining sufficient protein solubility.[29,50] The most substituted peptide, 18A, only contains four conserved residues from GCN4, and thus only the base-specific interactions are conserved, while most all of the electrostatic interactions are removed. 4A and 11A have fewer alanine substitutions than 18A, so they can contact to the DNA by more base specific interactions as well as more non-specific electrostatic interactions with the DNA backbone. 4A, 11A and 18A all retain their DNA binding function; but the DNA binding affinities of 4A and 11A are about 7-fold lower compared to GCN4, and the DNA binding affinity of 18A is 1000-fold lower. This is thought to be due to the loss of the electrostatic contacts with DNA. Increasing the alanine content also increases the protein α-helicity; the GCN4 basic region is not highly α-helical until selective binding to the major groove of target DNA. Therefore peptide with more alanine content is closer in form to the active conformation of the GCN4 basic region. 11A retains specific and non-specific interactions with DNA and its α-helicity approaches that of the 18A peptide, so 11A appears to be the best balance between

α-helical preorganisation and required electrostatic interactions with DNA. All the helicity was characterised by circular dichroism. The GCN4-DNA binding is enthalpically driven, but is unfavourable entropically.[54] 18A is a preorganised helix and therefore will not pay the folding penalty (entropy) on binding to DNA. However, at the same time the electrostatic contribution (enthalpy) has also been reduced. The 1000-fold lower DNA binding affinity of 18A therefore indicates that the preorganisation of peptide structure is not sufficient to compensate for the loss of favourable electrostatic interactions. It was therefore concluded that the electrostatic interactions (enthalpy) are more important than preorganisation (entropy) in determining the GCN4-DNA binding affinity.[50,51]

1.3.3 Moving the basic region towards the C-terminus of GCN4

Oakley and co-workers reversed the GCN4 basic domain and leucine zipper domain in sequence but the residues in the two subdomains are unchanged, and the dimer binds to DNA site contains the inverted CRE DNA half site with GCN4 affinity.

They synthesised a GCN4-56 variant with the basic region located towards the C-terminal instead of the N-terminus where it is located in wild type GCN4.[34] In their design, the basic DNA binding region containing residues 251-281 towards the N terminus was moved the C terminus (Table 2); a 4 amino sequence between the basic region and the leucine zipper was introduced as a hinge which would provide the required 3-dimensional relationship between the dimerisation interface, the DNA interacting residues and the exposed bases on the target DNA. The coiled coil leucine zipper was placed at the N terminus.

Table 2 The wild GCN4-56 sequence and rGCN4, in which residues 251-281 towards the N terminus has been moved the C terminus.

GCN4-56				
	N-Cap	**Basic region**	**Hinge**	**Leucine zipper**
MK	DPAAL	KRARNTEAARRSRARK	LQRV	KQLEDKVEELLSKNYHLENEVARLKKLVGER

rGCN4				
	Leucine zipper	**Hinge**	**Basic region**	**C-Cap**
M	KQLEDKVEELLSKNYHLENEVARLKKLVGE	LQRV	KRARNTEAARRSRARK	AALKG

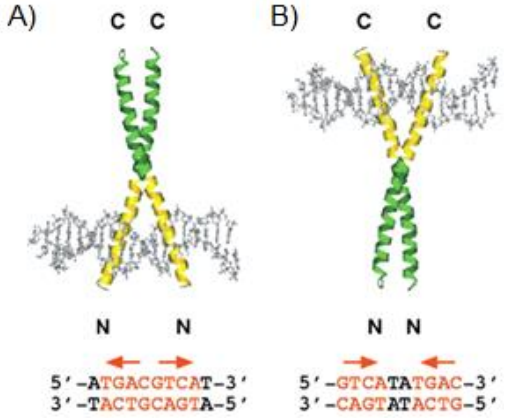
Because the correct spacing between the two subdomains in reversed GCN4 is unknown, the reversed GCN4 peptide was investigated with different hinge sequences (table 3). High affinity binding was only found for rGCN4-(+3) to the INV2 DNA site which contains the inverted half site based on CRE. rGCN4-(+3) contains a 7 amino acid hinge domain, corresponding to an integral number of a heptad repeat. Modest DNA binding affinity was only observed for rGCN4-(-1) which is related to rGCN-(+3) by close to a single turn of a helix. The worst binding were found for rGCN4-(+1) and rGCN4-(+4) which was again related by one turn of α-helix. Therefore when the insertion between the two subdomains contains 7 residues, which is equal to two turns of an α-helix, it maintains the position of the binding residues with the correct orientation with respect to the DNA target site.

Table 3 rGCN4 with various hinge domains were synthesised in order to establish the required space between the two subdomains, so as to retain their correct spatial relationship for DNA binding.[34]

rGCN4			
	N-terminus	**hinge**	**C-terminus**
rGCN4(-2)	Leucine Zipper	LV	Basic region
rGCN4(-1)	Leucine Zipper	LRV	Basic region
rGCN4(0)	Leucine Zipper	LQRV	Basic region
rGCN4(+1)	Leucine Zipper	LQRRV	Basic region
rGCN4(+2)	Leucine Zipper	LQLQRV	Basic region
rGCN4(+3)	Leucine Zipper	LQKLQRV	Basic region
rGCN4(+4)	Leucine Zipper	LQKQLQRV	Basic region

Because rGCN4 is reversed, its DNA binding site is different from that of wild-type GCN4-56. In the GCN4-56-CRE complex, the basic residues towards the N-terminal end contact base pairs on the outside of the DNA site, whereas basic residues near the leucine zipper domain contact the residues at the center of the DNA site. In rGCN4, basic residues which were near to the N-terminus in GCN4-56 are now placed towards the leucine zipper domain, while the residues which were near to the leucine zipper domain in GCN4-56 are now placed next to the N-terminus; and thus the DNA half site should be inverted to maintain the selective DNA contacts (Figure 10).[34]

Figure 10 A) GCN4-56 binding to CRE DNA and B) rGCN4(+3) binding to INV2 DNA. The leucine zipper is in shown green and the basic region in yellow, figure adapted from reference 34 with permission from Biochemistry. Copyright (2001) American Chemical Society.



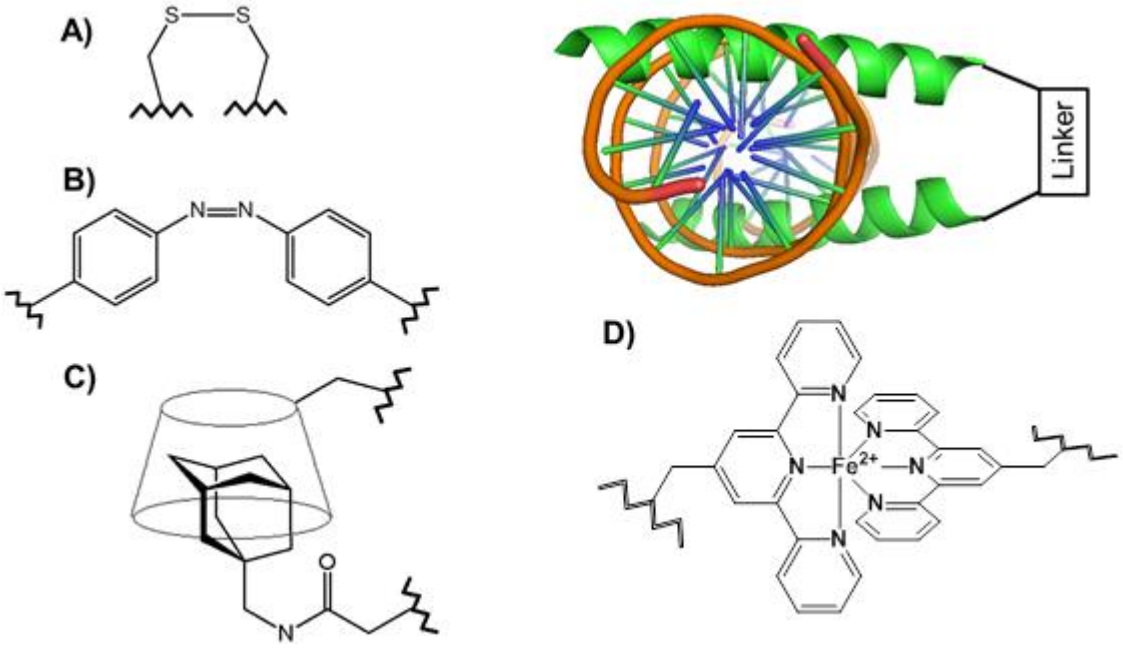
1.4 Artificial DNA binding peptide dimers based on GCN4

GCN4's DNA binding is highly dependent on dimerisation. This was illustrated by a study which showed that short basic region peptides can bind to DNA with similar affinity to native GCN4, when they are covalently linked by a flexible disulfide bond, see Figure 11A.[35] Various related peptide dimers inspired by this study have subsequently been synthesised and investigated for DNA binding, see Figure 11.[35-38]

Kim and co-workers dimerised a short peptide corresponding to the basic region of

GCN4 (amino acids 222-252 of GCN4) by the formation of a flexible disulfide linker, through an introduced Gly-Gly-Cys sequence towards the C-terminus (see Figure 11A). Their results revealed that the basic region of GCN4 is sufficiently long for sequence-selective DNA binding and that dimerisation is the primary function of the leucine zipper domain.[35]

Figure 11 Figure summarizing the reported linkers used as artificial dimerisation domains of peptides based on the basic region of GCN4 or MyoB for DNA binding. A) disulfide bridge.[35] B) azobenzene linker as photo switch,[36] C) a β cyclodextrin-adamantyl complex linker[37] and D) an bis(terpyride) iron complex.[38]



1.4.1 Synthetic peptide dimer with an azobenzene linker based on GCN4

Mascarenas and co-worker synthesised a DNA binding peptide dimer based on basic region of GCN4, with a light-modulated azobenzene linker, see Figure 11B.[36] In their design, the linker molecule is coupled to the peptide through the Cys side chain of an additional Gly-Gly-Cys sequence at the C terminus. The azobenzene compound was synthesised and coupled with the peptide as the trans conformation. Irradiation of an aqueous solution of the trans compound at 365 nm yielded a mixture

of cis:trans (95:5) isomers due to a light activated trans to cis conformational rearrangement. The cis isomer binds to DNA containing target site, with 60-70 times more efficiency than the trans isomer. The trans to cis isomerisation can also be carried out in the presence of DNA, but was found to be eight times slower than in the absence of DNA.

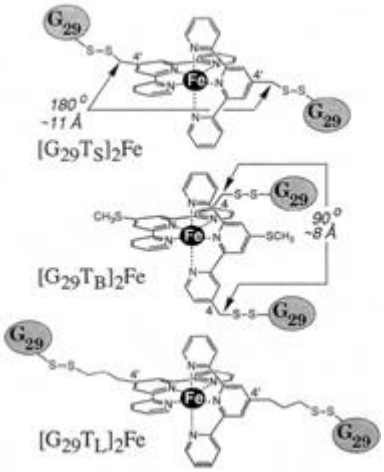
1.4.2 Synthetic peptide hetero dimer with β-cyclodextrin-adamantyl complex linker

Morri and co-workers reported a peptide dimerised using a non-covalent linker which binds DNA containing the CRE site, see Figure 11C.[37] The 23 amino acid peptide sequence was based on the basic domain of GCN4. The peptide sequence with C-terminal β-cyclodextrin (G23-CD) and a peptide with a C-terminal adamantly group (G23-AD) form a 1:1 heterodimer. G23-AD cannot bind to CRE DNA, while G23-CD appeared to bind to CRE DNA as a homo dimer, but with reduced affinity than the G23-AD/G23-CD mixture. The binding of the heterodimer was demonstrated by gel electrophoresis and circular dichroism.[37]

1.4.3 Synthetic peptide dimer with a bis(terpyride) iron (II) complex linker

Schepartz and co-workers prepared a series of DNA binding peptide dimers which are composed of two 29 amino acid peptides based on the basic region of GCN4, joined by disulfide bonds to 2 different positions of a bis(terpyride) iron(II) complex, resulting is systems with differing distances between the two DNA binding peptide domains (Figure 11D and12).[38]

Figure 12 The peptide dimer with iron complex linker prepared by Sdepartz and co-workers, figure is adapted from reference 38. Copyright (1993) the National Academy of Sciences.



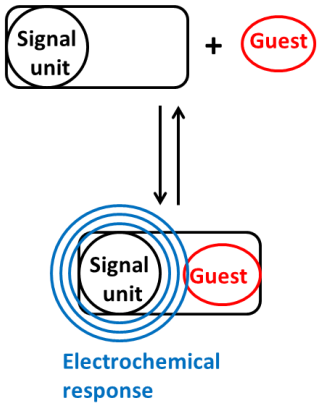
These dimers display different affinities for GCN4 DNA target sites. $[G_{29}T_S]_2Fe$ binds to the CRE site, but no CRE binding occurs with $[G_{29}T_L]_2Fe$ or $[G_{29}T_B]_2Fe$. Natural GCN4 binds to the CRE and AP1 sites with comparable affinity, whereas $[G_{29}T_S]_2Fe$ binds to CRE site with a 150-fold preference over AP1. The result is that a synthetic molecule could be the designed with a greater sequence specificity than GCN4 itself. [38]

1.5 Ferrocene Complexes as Electrochemical Sensors

Ferrocene complexes have been widely used in the electro-analytical sensors because of ferrocene's attractive electrochemical properties, including a reversible oxidation and reduction and stability in aqueous conditions.[55] In this section, ferrocene complexes designed as sensors to detect different analytes will be introduced. How the sensors interact with analytes in order to trigger an electrochemical response (modulation) in the solution phase, or immobilised on an electrode surface, as well as the sensors' electrochemical signal loss for the detection of analyte, will be introduced. The basic principal behind an electrochemical

sensor, involves the change in the electrochemical output of a host molecule, which consists of a binding site for guest binding and a reporter group in close proximity, see Figure 13.[56] This section focuses on sensors in which the reporter group is ferrocene. The two major approaches for surface immobilised ferrocene biosensor design, will be introduced.

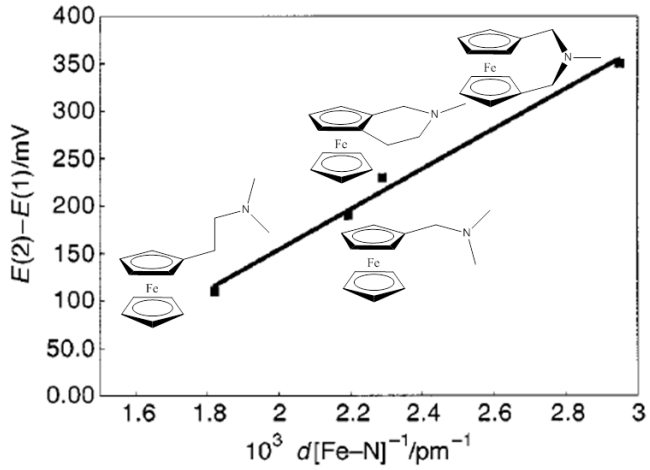
Figure 13 The binding of the guest species causes a change in the electrochemical output.



1.5.1 Ferrocene complexes as cation and anion sensors

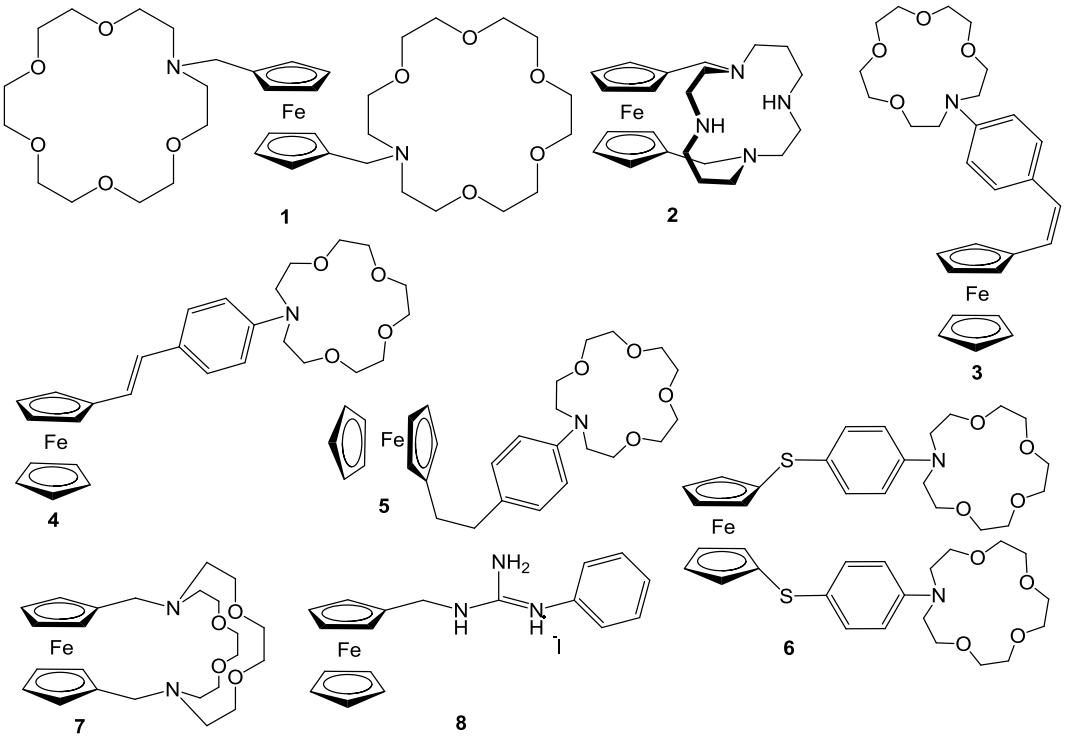
Plenio and Diodone studied more than twenty ferrocene nitrogen compounds with which a proton acts as a guest cation, and they reported a linear relationship between the anodic shift in potential and the distance between the proton and ferrocene, see Figure 14.[57] The ferrocene compounds are more difficult to oxidise when protonated, resulting a higher potential of the ferrocene-ferrocenium redox couple. The through space electrostatic interaction between the ferrocene and the protonated nitrogen, causes the potential to shift linearly against the inverse distance.

Figure 14 The anodic potential shifts plot as a function of inverse Fe-N distances.



Beer and co-workers demonstrated that the single redox active ferrocene receptor can also detect two different cations.[58,59] They found that molecule **1** (Figure 15) containing two metal binding sites, can sense the presence of barium, magnesium or a mixture of the two. There are three different anodic potential shifts of 150, 395 and 275 mV respectively for the three complexes: with $2Ba^{2+}$, $2Mg^{2+}$ and $Ba^{2+}{\cdot}Mg^{2+}$.[58,59]

Figure 15 Structures of the small ferrocene complex sense ions.



Plenio and Aberle prepared a ferrocene bridged cyclam **2** which display different large redox potential shifts, by the through space electrostatic interactions in the presence of different transition metal cations including Co^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+}.[60-63]

Beer and co-workers prepared some redox responsive crown ether molecules**.** These are consisted of a crown ether unit and a conjugated linked ferrocene unit (**3**, **4)**, as well as a saturated analogue (**5**).[64] The addition of magnesium cations resulted in anodic shifts in the ferrocene oxidation wave of the three molecules in cyclic voltammetry. The addition of Mg^{2+} into the conjugated receptor systems in which the heteroatoms in the crown ether unit are linked to the redox center by a π-electron system, causes the anodic potential shifts of 100 and 110 mV for the ferrocene oxidation wave of **3** and **4** respectively. For the saturated analogue **5**, the addition of Mg^{2+} only shifts the anodic peak potential less than 20 mV, suggesting that the electrochemical recognition mechanism in molecules **3** and **4** are majorly through the bond.

The conformational change induced by the guest binding may also provide a mechanism for changing the electrochemical behavior of the redox active center.[65]

The addition of K^+ cations into the solution of the sulfur linked bis crown ferrocene (**6**) causes an unusual cathodic potential shift of 60 mV.[66] This cathodic potential change is not due to the interactions through space or through bond interactions. It forms a 1:1 complex with K^+ cations, sandwiching the cation between the two crown ether units. The complexation causes a reorientation of the sulfur lone pairs towards the iron center of the ferrocene, thereby increasing the electron density on the iron, resulting in a cathodic shift of the ferrocene redox potential.

A direct interaction between the ferrocene unit and a cation can also produce a

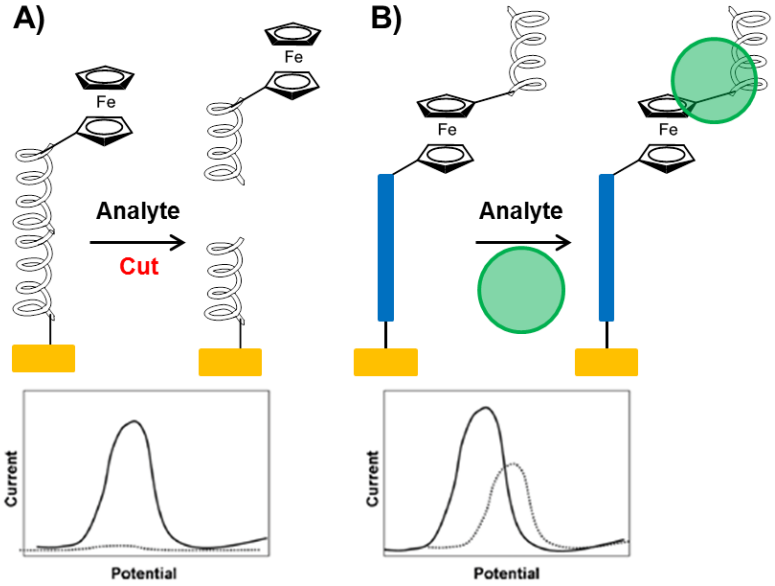
change in the redox property of the ferrocene unit. **7** can form the direct co-ordination with silver cations, to produce a $\Delta E_{1/2}$ of 282 mV larger than that observed with Ca^{2+} cation, and this is against its smaller charge to size ratio of the silver cation.[67] The shift in $\Delta E_{1/2}$ of ferrocene unit by Ag^{+} is much larger than predicted due to its charge to size ratio. The x-ray crystal structures of the resulting Ag^{+} and Na^{+} complexes, indicate the Ag to Fe distance in the Ag^{+} complex is 3.37 Å and shorter than the distance of 4.39 Å between Na and Fe. UV-vis spectra reveal that it is because the ferrocene group acts as an electron donor to the Ag^{+} cation.[67]

Anionic species is challenging to be detected by electrochemical sensors. The binding of anions are more difficult with respect to cations because of their pH sensitivity and geometry range. Beer and co-workers prepared ferrocene containing molecule **8**, and it can bind anions. The presence of $H_2PO_4^{-}$ causes a cathodic shift of redox potential in ferrocene by the through space interactions for 125 mV and 70 mV in DMSO and methanol-water solution respectively.[68]

1.5.2 Ferrocene peptide derivatives as biomolecule sensors

There are two major approaches for the surface immobilised ferrocene peptide sensors. The first approach is the ‘signal off’ system in which the presence of the analyte will cause a drop in current. The second approach is the modulation system, in which complexation of the peptide with the biomolecule analyte in solution, produces an increase in the redox electrochemical response which is proportional to the analyte concentration, see Figure 16.

Figure 16 Two different approaches for the detection of a biological analytes using surface immobilised ferrocene peptides. A) 'Signal off' approach, in which the loss of the ferrocene tag cause a decrease in the current. B) Modulation approach, in which the complexation of the ferrocene peptide with the analyte causes an electrochemical response at a different potential. Darker line is the signal in the absence of analyte, and grey line is that in the presence of the analyte.[55]



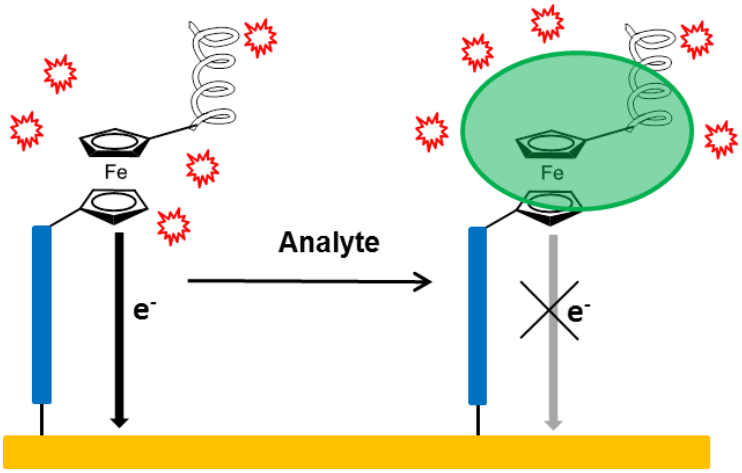
A ferrocene tag attached to the N-terminus of a Gly-Asp-Gly-Asp-Glu-Val-Asp-Gly-Cys peptide sequence immobilised on a gold electrode surface can detect the caspase activity which occurs in apotosis.[69] This system was therefore used to detect the presence of the caspase-3 enzyme.[69] Incubation of the ferrocene peptide immobilised electrode in apoptotic cell lysates causes a significant decrease in the ferrocene redox current due to the recognition and cleavage of the N-terminal Asp-Glu-Val-Asp peptide sequence by caspase-3. The loss of a ferrocene tag from the sensor results in a decrease in the current.

The 'signal off' approach has also been used for the detection of prostate specific antigen (PSA).[70] PSA is a serine protease used in the early detection of prostate cancer.[71,72] The ferrocene peptide sequence of Cys-His-Ser-Ser-Leu-Lys-Gln-Lys-ferrocene, immobilised on a gold electrode surface, was used for the detection of PSA. PSA hydrolyses the Ser-Ser bond in the peptide sequence to cleave the

ferrocene tag and cause a decrease in the redox current. The decrease in current percentage is proportional to the concentration of PSA in the range of 0.5 to 40 ng mL^{-1}.[70]

The modulation sensor approach has been used for the detection of HIV-1 protease (PR).[73] The gold surface attached ferrocene conjugate of pepstatin (Val-Cal-Sta-Ala-Sta) with two residues of 3*S*, 4*S*-4amino-3-hydroxy-6-methylheptanioc acid, was used because pepstatin can bind to the active sites of most aspartic proteases including HIV-1 PR.[74] The interaction between the conjugate pepstatin and HIV-1 PR, was detected by cyclic voltammetry. An increasing concentration of HIV-1 PR increases the formal potential and decreases the current intensity of the ferrocene peptide conjugate immobilised on the surface, therefore the encapsulation of the ferrocene group by the protease makes the oxidation of the ferrocene group more difficult due to the shielding of counter ions access to the ferrocene group, see Figure 17. A linear relationship between HIV-1 PR concentration (up 100 nM) and current, was reported.[74]

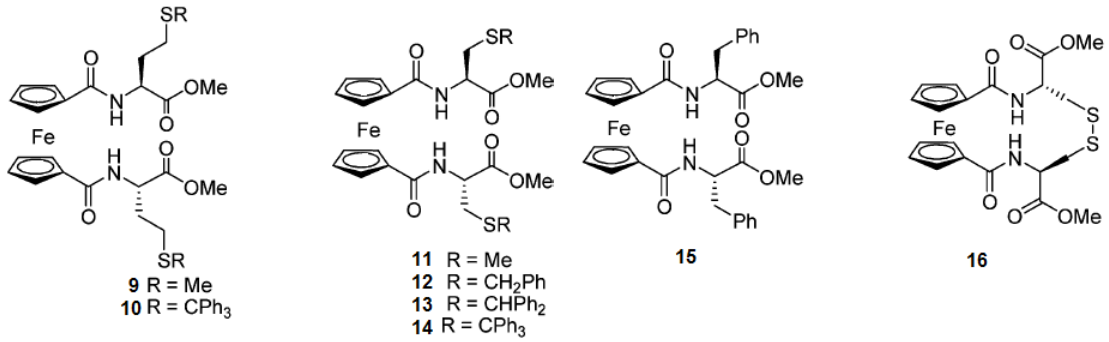
Figure 17 Counter ions (red) in solution experience restricted access to the ferrocene group when the ferrocene peptide conjugate is encapsulated by HIV-1 PR (green).[55,73]



1.5.3 Ferrocene peptide conjugates as ion sensors

The complexation of ferrocene peptide conjugates with metal ions generally involves the backbone coordination and N, O chelation, or interactions with amino acid side chains.[55] For example, the ferrocene dipeptides incorporating sulfur side chains and the sulfur free analogues (**9–16**), were studied by electrochemistry in the presence and absence of various heavy metal ions in solution.[75] Some of these ferrocene peptides incorporated protected Met or Cys amino acids. The sulfur-containing amino acids are chelators of heavy metal ions, such as Hg^{2+}.[76] The titration of heavy metal ions into the ferrocene peptide solutions produced an anodic shifts of 30 to 70 mV in the redox potentials of **9** to **14**. The addition of Hg^{2+} produced higher redox potential shifts than Cd^{2+}, Pd^{2+}, Zn^{2+} and Ag^{+}. NMR studies revealed that the interaction with sulfur containing side chains is more important than with backbone amide groups. Compound **15** is a sulfur free analogue, for which there is no significant electrochemical change on addition of Hg^{2+}, supporting that sulfur-Hg coordination is required for a significant anodic potential shift. Compound **16** which contains a disulfide bond, is a known host for group II metal ions, however it does not undergo a similar electrochemical response on the addition of transition metal ions at the concentration of 0-60 ppm, and the reason is not clear.[76]

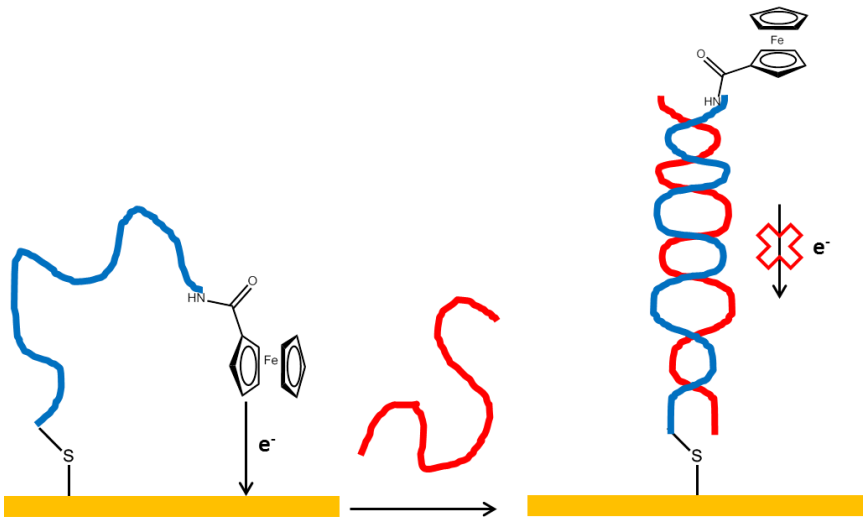
Figure 18 Structures of the ferrocene peptide conjugate as ion sensor.



1.5.4 Ferrocene peptide nucleic acid (PNA) and ferrocene deoxyribonucleic acid DNA conjugates

Peptide nucleic acid (PNA) is an analogue of deoxyribonucleic acid DNA and it is composed of *N*-(2-aminoethyl) glycine units linked by peptide bonds with purine and pyrimidine nucleobases on side chains.[77] The PNA/DNA hydrid is similar to double stranded DNA with a similar H-bonding network, but it does not experience the same interstrand electrostatic repulsion. A ferrocene PNA monomer immobilised on electrode surface for the detection of a complementary sequence strand of DNA has been developed by Metzler-Nolte. A PNA sequence was prepared with Cys on the N-terminus, for surface immobilisation, and a ferrocene group at the C-terminus as a redox reporter.[78] The ferrocene PNA experiences an anodic potential shift of 440 mV on hybridisation with complementary DNA, due to the distance dependent charge transfer rate of the sensor system, see Figure 19.[79] The increasing distance between the ferrocene redox reporter and the surface causes an anodic shift of the peak potential and a decrease in the peak current of the ferrocene reporter.[80]

Figure 19 The hybridisation of ferrocene PNA with complementary DNA increases the distance between the ferrocene group and the electrode surface and make the ferrocene group harder for electrochemical access.



Kwak detected DNA hybridisation by the cleavage of S1 nuclease which only cleaves single strand DNA. Ferrocene DNA immobilised on a gold electrode hybridised with complementary and noncomplementary DNAs, and then the nonhybridised DNA was cleaved by S1 nuclease. CV and DPV of the dsDNA detected the differences between the three target hybridisation on the electrode.[81]

Hsing and co-workers provided a method to detect sequence specific DNA without immobilizing the probe molecule onto a solid support.[82] In the method, ferrocene labelled PNA hybridised with a strand of complementary DNA directly in solution. The negatively charged hybrid and the neutrally charged ferrocene labelled PNA were studied by differential pulse voltammetry (DPV) with a negatively charged indium tin oxide electrode (ITO) and positive charged poly (allylamine hydrochloride) (PAH) layer coated ITO electrode. Due to the electrostatic repulsion between the hybrid and the ITO electrode, the hybrid has a smaller signal than the free ferrocene-PNA monomer as measured by DPV. Due to the electrostatic attraction between the hybrid and PAH-ITO electrode, the hybrid displays a stronger signal than the ferrocene-PNA probe by DPV, and the signal is proportional to the amount of complementary DNA with respect to the ferrocene-PNA.

Jenkins labelled a single strand oligonucleotide with ferrocene and characterised by DPV prior to and after the cleavage by a nuclease.[83] The ferrocene-oligonucleotide displays a weaker signal in DPV prior to the cleavage in solution, as the oligonucleotide hinders the migration of ferrocene to the electrode.

1.6 Aims of the project

GCN4 is a bZip peptide that has been widely studied. The replacement of the leucine zipper domain by chemical linkage of the basic domains will retain its DNA binding ability. Ferrocene tagged small molecules have been widely studied for ion sensory applications, while ferrocene tagged biomolecules have been widely studied for ion and biomolecule sensory applications.

This project utilises the flexible ferrocene (due to free rotation of the Cp rings) as the dimerisation unit for two short GCN4 basic domains. The combination of the characteristic and attractive electrochemical properties of ferrocene and the DNA binding ability of the dimerised basic domains should provide an electrochemically active ferrocene peptide dimer, which would potentially bind to DNA.

Various DNA binding peptide dimers with various sequences based on native GCN4 with artificial dimerisation linkers have been reported. The hinge domain maintains the spatial relationship between basic domain and leucine zipper for DNA binding. In this project, the GCN4 basic domains are dimerised with ferrocene, which has not been reported in the literature. The optimal conformation for correct alignment of the two α-helical basic domains for effective DNA target site binding, was studied by the ferrocene peptide dimers with various hinge sequences. This will also be used to investigate the shortest hinge domain to maintain the spatial relationship between the basic domain and the ferrocene unit for DNA binding.

The ferrocene peptides will be studied in solution rather than their being attached to an electrode surface, and thus neither the peptide chain nor the DNA will play a direct role in the electron transfer pathway. Therefore, if the ferrocene peptides bind to the DNA target site, the new environment experienced by the ferrocene group,

including the surrounding charges and steric hindrance, may play an important role in the electrochemical response. Changes to the ferrocene peptide, in the presence and absence of both non-specific and target DNA, will be investigated by electrochemical methods including cyclic voltammetry and differential pulse voltammetry. The various ferrocene peptide dimers, with different hinge sequences, will be compared and any differences in their electrochemical response will be discussed in terms of the DNA phosphate backbone-ferrocene distance established from the molecular dynamics simulation models. This is because the negatively charged DNA backbone may have a significant effect on the ferrocene by through space electrostatic forces. In contrast, how the ferrocene peptide dimers interact with non-specific DNA is unknown, and thus the electrochemical behavior of the ferrocene peptides in the presence of non-specific DNA will be studied and compared to that with non-specific DNA.

Chapter 2 Experimental techniques

2.1 Introduction

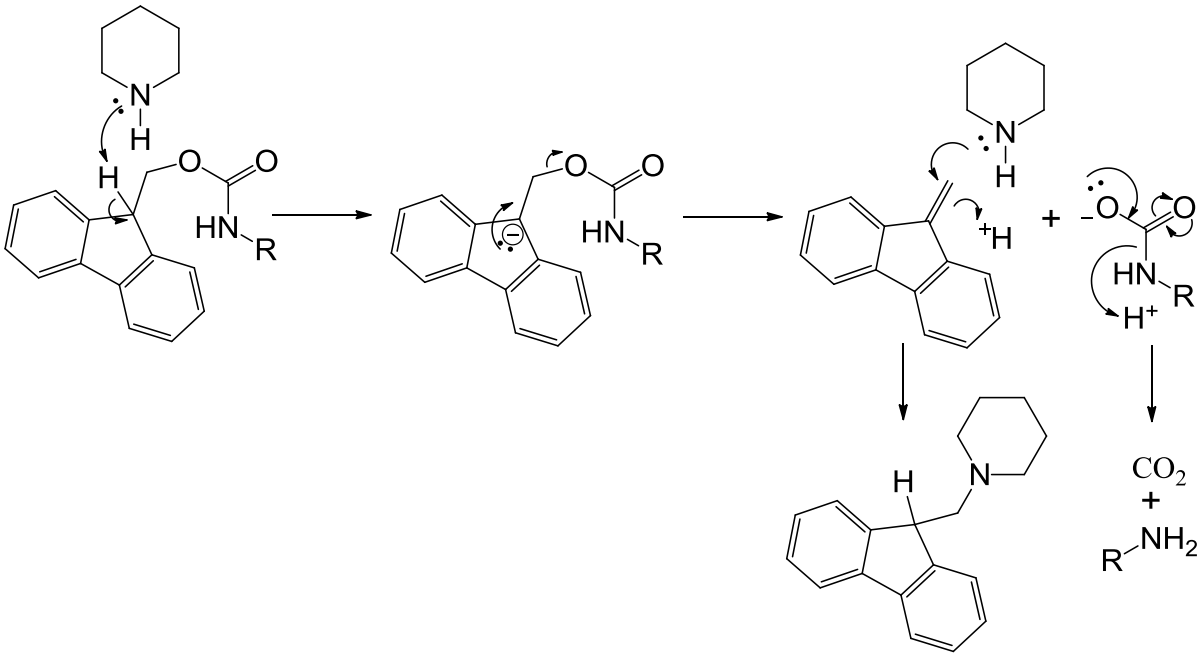
In this chapter, the experimental techniques used in this project will be introduced. The preparation of side chain protected and side chain deprotected peptides inspired by the bacterial transcription factor GCN4, was carried out by microwave-assisted solid phase peptide synthesis (SPPS). The peptides were purified by preparative HPLC, characterised by MS and their purity is evaluated using analytical HPLC. All low molecular weight ferrocene complexes synthesised were characterised by ^{1}H and ^{13}C NMR, UV-visible, CV and MS. Ferrocene peptide dimers were subsequently synthesised and again characterised and purified by HPLC and MS and quantified by UV-visible spectroscopy. The GCN4 basic domain, the peptide adopted in this study, is unstructured in the absence of DNA containing a target site,[33] and this therefore provides us with a method by which to study binding of ferrocene peptide dimers to various DNA sequences. The folding of peptides and ferrocene peptide dimers was therefore measured using circular dichroism (CD) spectroscopy. Ultimately all the ferrocene peptide dimers (44AA, 48AA, 50AA, 52AA and 56AA) were studied by cyclic voltammetry (CV) in order to study the redox character of the ferrocene group. Differential pulse voltammetry (DPV) has higher sensitivity than CV and so this technique was adopted to characterise the sample at lower concentrations.

2.2 Fmoc solid-phase peptide synthesis (SPPS)

Solid-phase peptide synthesis (SPPS) was developed by Merrifield in 1963.[84] SPPS involves the use of a support resin in peptide synthesis. The first amino acid is coupled to the resin. The majority of amino acid side chains are protected with temporary protecting groups, to prevent unwanted coupling and modifications to

these amino acid side chains. Protecting groups can be selected which are stable under the peptide synthesis conditions. However it is also possible to use alternative protecting groups which can be selectively removed, offering exciting opportunities in peptide synthesis. Similarly the N-terminus of the amino acid is protected with a fluorenylmethyloxycarbonyl (Fmoc) protecting group in order to prevent unwanted side reactions. Fmoc is an acid-stable protecting group used at the N-terminus and can tolerate, for example the TFA cocktail used for the cleavage of amino acid side chain protecting groups. It is stable to tertiary amines like *N,N*-diisopropylethylamine and pyridine. Therefore Fmoc is selectively removed by base induced β-elimination, see Scheme 1.

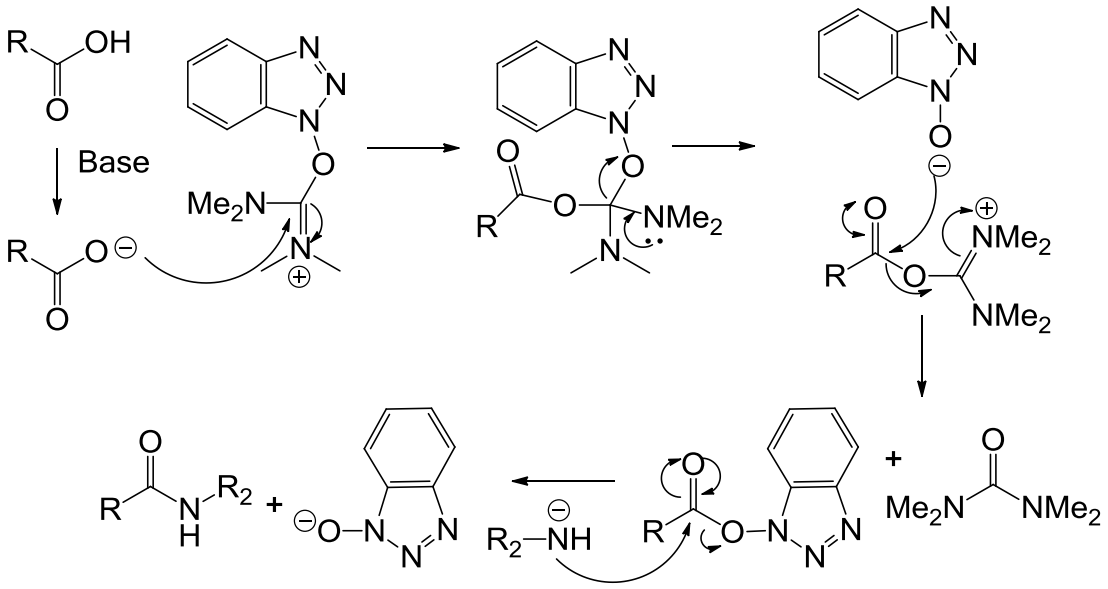
Scheme 1 Mechanism of Fmoc cleavage using piperidine. R=peptide on resin



Having removed the Fmoc protecting group, peptide bonds are formed between the free amine (N-terminus) and the carboxylate (C-terminus) of a new Fmoc protected

amino acid using HBTU. Scheme 2 shows the mechanism of the activation of the carboxylate using HBTU.

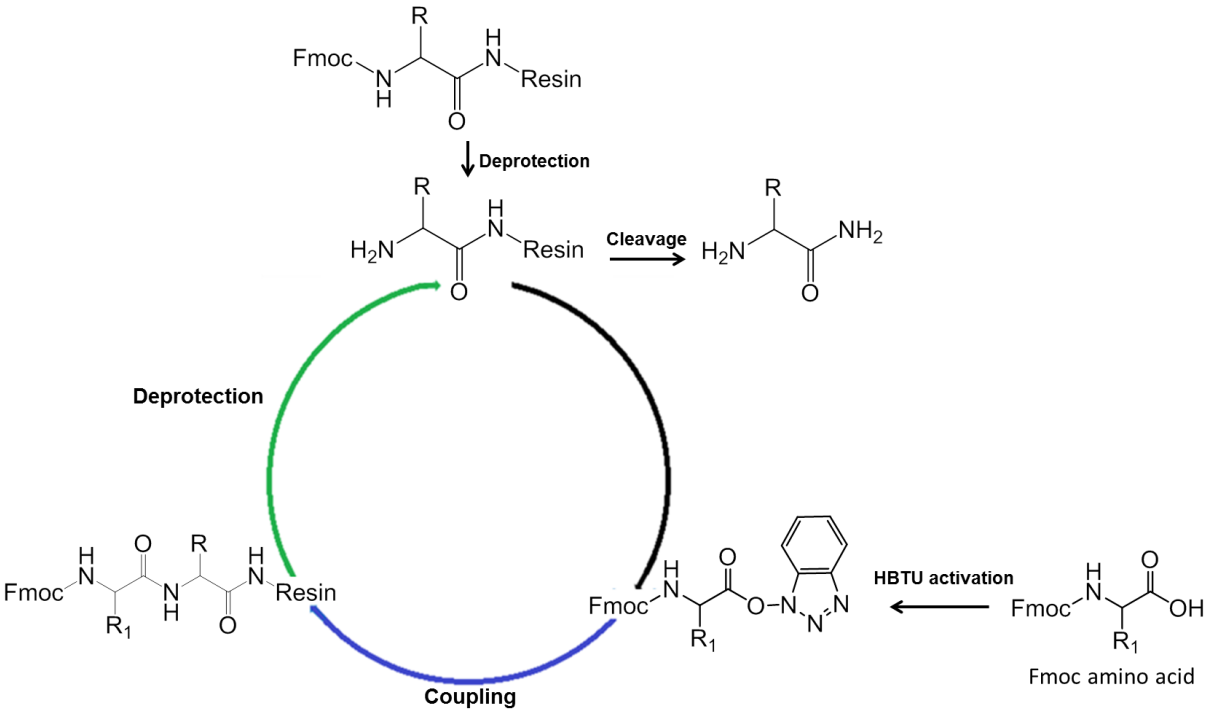
Scheme 2 Mechanism of peptide bond formation via HBTU activation. R = the peptide on resin, R_2= the new amino acid.



The single amino acid loaded resin is washed and filtered, to remove any unreacted reagents and by-products. The Fmoc group located at the N-terminus of the amino acid is removed by piperidine and N,N-diisopropylethylamine in DMF solution (20/20/60), the resin again washed and filtered before coupling of the next amino acid. The amino acid cycles are repeated until the sequence is finished, thereby sequentially growing the peptide from the C- to the N- terminus. All the side chain protecting groups and the resin can be cleaved from the peptide by TFA/thioanisole/1, 2-ethanedithiol (EDT)/anisole (90/5/3/2), which includes the cleavage agent TFA and the nucleophilic scavengers thioanisole, EDT and anisole, see Figure 1. Alternatively a resin such as 2-chlorotrityl chloride resin can be used which allows for the cleavage of the peptide from the resin under mildly acidic conditions such as 20% 2,2,2-

trifluoroethanol (TFE), 20% acetic acid and 60% DCM, whilst retaining all the amino acid protection groups.

Figure 1 Solid-phase peptide synthesis cycle. The first Fmoc group is deprotected from the resin bound amino acid, which subsequently couples with the HBTU activated second Fmoc amino acid building block, thereby extending the peptide sequence on resin. The Fmoc is again deprotected in preparation for the next synthesis cycle or for cleavage of the peptide from the resin.



Microwave assisted SPPS shortens the reaction times as the microwave energy activates molecules with a dipole moment and allows for rapid heating at the molecular level.[85] The standard (25 Watt, 75°C, 300 seconds) coupling method was applied for most amino acid coupling. However coupling of Arg was carried out at room temperature, due to the formation of γ-lactam by the Fmoc-Arg(Pbf)-OH building block during carboxyl activation with microwave heating.[86] The coupling of cysteine was performed at the lower temperature of 50°C,[87] due to potential racemisation of the cysteine at the higher temperatures obtained during microwave

heating. This would also apply to histidine, however this residue was not present in our peptide sequences.[87]

2.3 Ultraviolet-visible (UV-VIS) spectroscopy

Many organic molecules and functional groups are transparent in the ultraviolet (UV) and visible (VIS) regions of the electromagnetic spectrum, which spans between 190 nm and 800 nm. The absorption of specific wavelengths of light causes gaps in the spectrum, resulting in what is called an absorption spectrum. The absorption spectrum in this range can provide useful information about conjugated π-electron systems. The energy of the radiation absorbed is the same as the energy difference between the low and high energy states.[88] The relationship between absorption, concentration, path length and the ability of a species to absorb light at a specific wavelength (molar absorptivity) is described by the Beer-Lambert Law.[88] This formula is used widely for the quantification of the material.

$$A = \varepsilon c l$$

A = absorbance

ε = molar absorptivity

c = solute molar concentration

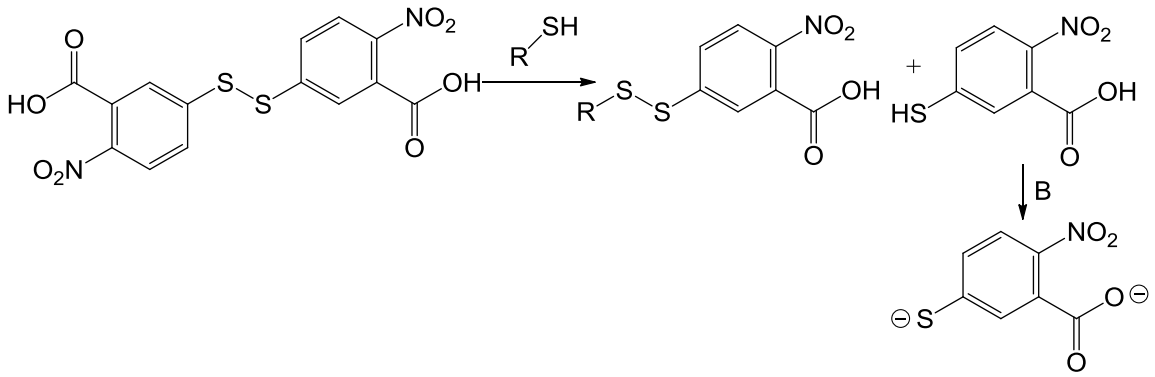
l = sample cell length in cm

2.3.1 The Ellman's test

Though peptides absorb in the 190-230 nm region due to the peptide bond, chromophores are used to quantify peptide concentrations. Alternative chromophores include tryptophan (ε_{280nm}= 5600 $M^{-1}cm^{-1}$), tyrosine (ε_{275nm}=1420 $M^{-1}cm^{-1}$) and phenylalanine (ε_{257}nm=179 $M^{-1}cm^{-1}$). The Ellman's test is used for the

quantification of available thiols concentration in solution by UV-vis spectroscopy.[89] In the Ellman's test, 5,5'-dithiobis-(2-nitrobenzoic acid) reacts with thiols to cleave the disulfide bond and results in the rapid formation of 2-nitro-5-thiobenzoate (NTB^-), which ionises to the yellow coloured NTB^{2-} dianion under neutral and basic aqueous conditions. The absorbance of NTB^{2-} at 412 nm and its extinction coefficient of 13,700 M^{-1} cm^{-1}, is used to establish the concentration of available Cys in the sample and thereby the concentration of peptide.[90]

Scheme 3 The reaction of a thiol with 5,5'-dithiobis-(2-nitrobenzoic acid) to yield NTB^{2-}.



2.4 Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy is an important technique with which to study the magnetic properties of various nuclei which can absorb and re-emit electromagnetic radiation when in a magnetic field. NMR spectra can provide information about the number of magnetically distinct environments for each atom being studied, as well as information about their environment.[88] Hydrogen (^1H) and carbon (^{13}C) are the most common nuclei routinely studied in NMR, and they are also the ones used in this project. NMR experiments performed on the low molecular weight ferrocene

complexes were all recorded in either D_2O or a deuterated organic solvent. All ^{1}H NMR spectra were recorded at 300 MHz on a Bucker AV III 300 NMR and all ^{13}C NMR spectra were recorded at 100 MHz on a Bucker AV III 400 NMR spectrometer. 1, 4-Dioxane was added in small quantities as an internal standard for ^{1}H NMR recorded in D_2O, whereas those collected in organic solvents were referenced to the solvent peak.[91] DEPT (Distortionless Enhancement by Polarisation Transfer) NMR was used to determine the multiplicity of carbon atom substitution with hydrogen. The ^{1}H NMR spectra were reported as follows: δ (ppm), the number of nuclei in the particular chemical environment (based on integration), the multiplicity of the signals (s: singlet; d: doublet; t: triplet; q: quartet; m: multiplet; dd: doublet of doublet; dt: doublet of triplet), coupling constants in Hz, followed by the assignment of the H in the structure. The ^{13}C NMR spectra were reported as follow: δ (ppm), the assignment of the C in the structure.

2.5 Mass spectrometry

Electron impact (EI) mass spectrometry (MS) is one of the most common MS techniques, providing information about the mass to charge ratio of ions in three steps. First, it streams high energy electrons to the molecules for bombardment (this ionisation method is called electron impact), and the ionised molecules are subsequently accelerated in an electric field. Second, it separates the accelerated ions according to their mass to charge ratio by traversing in a magnetic field. Finally, a detector can count a number of ions with particular mass to charge ratio. The output is then amplified and sent to the recorder which provides the mass spectrum.[33] Our samples were recorded by ESI or MALDI MS, which use different

ionisation techniques to EI-MS. The difference between the different techniques will be introduced in the following sub sections.

2.5.1 Electrospray ionisation mass spectrometry (ESI-MS)

In an ESI mass spectrometer, the analyte solution is dispersed as an aerosol. ESI-MS of peptides solutions results in the detection of multiple charged peptides. ESI-MS were recorded on a Waters Micromass LCT with direct infusion of peptide solutions prepared either in water/acetonitrile (50 % to 50 % with 1 % formic acid) or methanol. The cone voltage is 15 V and the capillary voltage is 3200 V. The source temperature is 130°C and the desolvation temperature is 300°C. The scan range is 200 to 2000 *m/z* and data is recorded in the positive ion mode. ESI-MS of the small molecular weight ferrocene complexes were recorded with direct infusion of methanol with the scan range between 200 to 1000 *m/z* and data is recorded in the positive ion mode.

2.5.2 Matrix-assisted laser desorption ionisation mass spectrometry (MALDI-MS)

In MALDI-MS, samples are prepared by mixing them with a matrix material and stabilizing them. Matrixes tend to be low molecular weight strongly absorbent compounds. In MALDI-MS, a laser ionises the matrix and causes the secondary ionisation of the sample. Sinapinic acid is typically used with our peptide samples; and gentisic acid for the low molecular weight ferrocene complexes. The reflectron grid reflects the ions towards the detector. All of our MALDI MS were recorded in positive mode on a Waters Micromass MALDI MX spectrometer. The laser

wavelength is at 337.1 nm with an average power of 3 mV at 10 Hz. All MALDI MS of peptide samples were recorded in linear mode with the range of 1000 to 10000 *m/z*. The MALDI MS of low molecular weight species were recorded in reflector mode with the range of 300 to 2000 *m/z*.

2.6 Reversed phase high-performance liquid chromatography (HPLC)

HPLC is a column chromatographic technique which separates the components of a mixture depending on their affinity for a solid and mobile phase respectively. A liquid sample or the sample solution, are forced to pass over a solid adsorbent material packed in a column using the flow of liquid solvent under high pressure. The molecules exiting the column are monitored by UV-vis of the solution at a particular wavelength, selected based on the absorbance profile of the desired molecule in the case of peptides this is 222 nm due to absorbance of the peptide bond. In reversed phase HPLC, the solid adsorbent material is silica modified to be non-polar by attaching hydrocarbon chains to its surface, which in a C18 column corresponds to an eighteen carbon chain, and a polar solvent is used as the mobile phase. Thus as the mixture passes through the column, there will be a strong attraction between the polar solvent and polar molecules whilst the polar molecules will not experience a strong attraction and thereby association, with the hydrocarbon chains attached to the silica of the stationary phase. In contrast, the non-polar components in the mixture will be less soluble in the polar solvent and tend to adsorb on to the non-polar stationary phase. Thus the polar molecules will pass through the column more quickly than the less polar molecules. A combination of the polarity of the molecules, the polarity of the mobile phase (which can be modified to include different solvents),

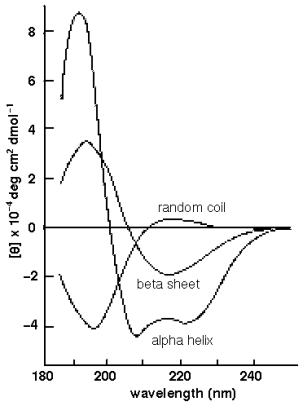
the nature of the stationary phase, contribute to the separation by HPLC. All our peptides are purified by preparative C12 column reverse phase HPLC as the peptide purity is evaluated by analytical C12 column reverse phase HPLC obtained using the same solvent gradient.

2.7 Circular dichroism (CD) spectroscopy

Circular dichroism (CD) is a spectroscopic technique, discovered by Aime Cotton and co-worker, which allows the study of differential absorption of left and right circular polarised light.[92, 93] This occurs when the chromophore is optically active, which is the case for peptides due to the chirality of amino acids and the protein fold. CD is therefore routinely used for the study of protein secondary structure.[94]

The CD for all α-helical peptides has two negative bands at 222 nm and 208 nm, and a positive band at 190 nm, see Figure 2. The band at 222 nm is related to the n→ π* transition of the amide bond. The bands at 208 nm and 190 nm are a result of the exciton splitting of the π → π* transition of the amide bond. The CD of β-sheet peptides have a negative band at 210 – 220 nm due to the n→ π* transition of amide, and a positive band at 198 nm due to the π → π* transition of amide, see Figure 2.[95]

Figure 2 CD spectra of an α-helical peptide, a β-sheet peptide and a random coiled peptide.[95]



The observed ellipticity in millidegrees is converted into molar residual ellipticity θ_{MRE} and reported in units of mdeg cm^2 $dmol^{-1}$ $residue^{-1}$. θ_{MRE} is calculated using

$$\Theta_{MRE} = [\Theta]/(10\times c\times d\times n)$$

Where:

c = concentration

d = cell path length in cm

n = number of residues in the peptide

The extent to which the peptide is folded into an α-helix can be calculated from the molar ellipticity at 222 nm using the following equation which uses the mean residue ellipticity of polylysine for 100% α-helix θ_{222nm} = -35,700 as a reference.[96]

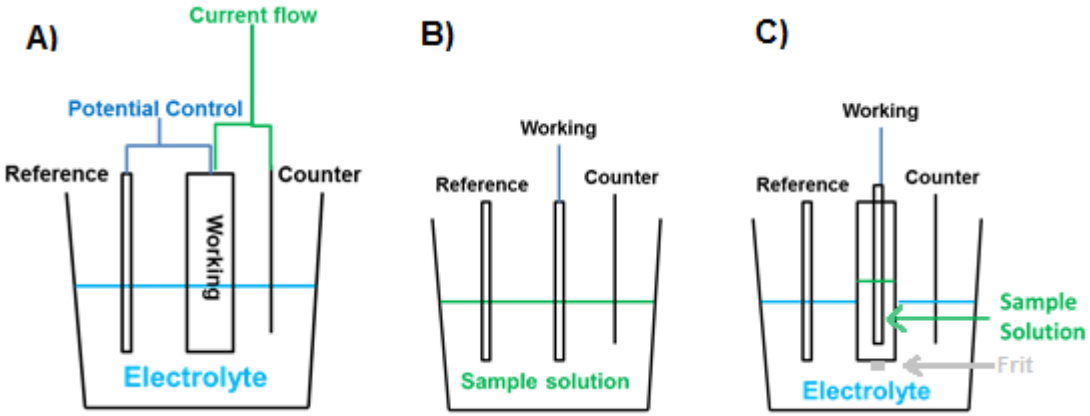
$$\text{Percentage folded as α-helix} = [\theta]_{222} \times 100/-35700$$

2.8 Electrochemistry

In voltammetry, the species of interest is analysed by the measurement of the current as a function of a potential which is varied. This provides a method for studying the redox potential of an electro-active system. The application of a potential can trigger a redox reaction, generating a current which can be measured. In our electrochemical studies, a three electrode cell is used which includes a working electrode, a reference electrode and a counter electrode, submerged in an electrolyte solution. The potentiostat connected to the three electrodes scan the potential at the working electrode. The potential is reported relative to that of the

Ag/AgCl reference electrode and the resulting current is monitored through the counter electrode, see Figure 3.

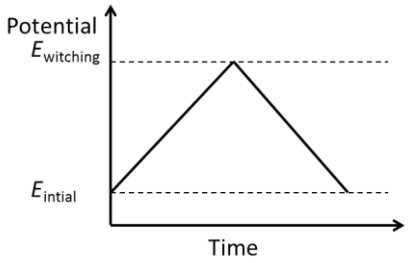
Figure 3 A) The three electrode system, The potential is applied between the working and reference electrodes, and the current flow is monitored between the working and counter electrodes. The two different experiment set-up for electrochemistry; B): all three electrodes are emerged in the sample solution; C): only the working electrode is in sample solution in sample holder, reference and counter electrodes are in the buffer solution. The frit is at the bottom of sample holder.



2.8.1 Cyclic voltammetry (CV)

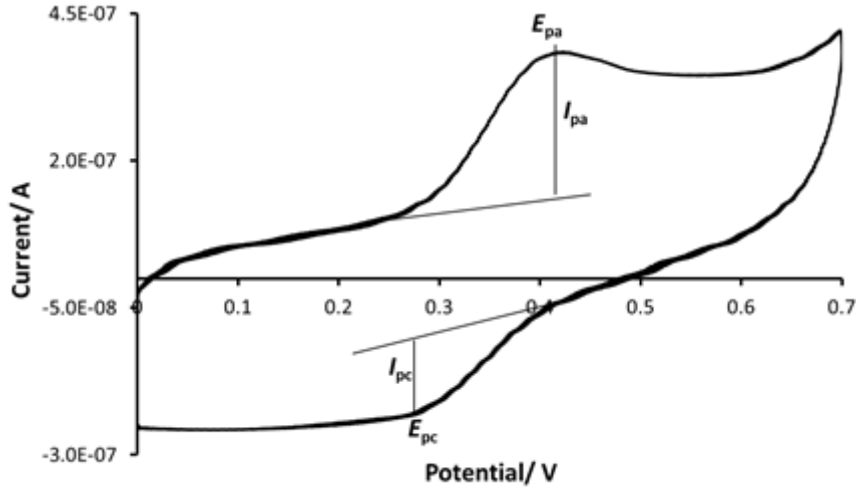
Cyclic Voltammetry (CV) is a potential sweep technique where the potential of the working electrode is linearly changed with a constant scan rate from an initial value, E_{intial}, to a switching value, $E_{switching}$, and then linearly reversed back to the original initial value, completing a full cycle (Figure 4).

Figure 4 Diagram showing the potential change as a function of time for one CV cycle.



The recorded current is plotted as a function of the potential, which for a reversible redox process results in a peak for the oxidation and reduction of the analyte in the forward scan and backward scan, respectively. The current versus potential diagram, contains a maximum i_{pa} (anodic peak current) at E_{pa} (anodic peak potential) in the anodic scan and a minimum current i_{pc} (cathodic peak current) at E_{pc} (cathodic peak potential) in the cathodic scan (Figure 5).

Figure 5 An example of a cyclic voltammogram of a non-reversible system, with ΔE_p more than 59 mV and $i_{pa}/i_{pc} > 1$



The cyclic voltammogram can provide important information concerning the reversibility of the system. In a fully reversible system, the heights of the forward and reverse current peaks are the same, and are separated by 59 mV for a single electron process for each molecule. These features should be independent of the scan rate if electron transfer is fast.[97] Slow electron transfer will cause a quasi-reversible electron transfer, see Figure 5.

The anodic and cathodic peaks in the CV are asymmetrical due to the polarisation overpotential to make the electron transfer process happen. Overpotential is the potential difference between a half-reaction determined reduction

potential and the potential in redox event observed in the experiment. It is the extra energy added to the reaction in order to drive the reaction to proceed. The overpotential is described in the Butler-Volmer equation:[89,99]

$$k_{ET} = k^{o} e^{-\alpha n F \eta / RT}$$

Where:

k_{ET} = rate constant of the electron transfer process

k^{o} = rate constant for standard heterogeneous electron transfer

α = transfer coefficient

n = number of electrons transferred for each molecule

R = universal gas constant

T = temperature in K

η = overpotential, and $\eta = E - E^{O'}$ ($E^{O'} = E_{pa} + E_{pc}/2$ = formal potential)

F = Faraday constant

The peak currents (i_{pa} and i_{pc}) are dependent on the scan rate, and are proportional to the square root of the scan rate, as described in the Cottrell equation.[100]

$$i = nFAC(\mathrm{v}D)^{1/2}(\pi)^{-1/2}$$

Where:

i = peak current (A)

n = number of electrons transferred per molecule

F = Faraday constant

A = surface area of the electrode (cm)

C = bulk concentration (mol/cm^3)

D = diffusion coefficient of diffusion species (cm^2/s)

v = scan rate (s^{-1})

2.8.2 Differential pulse voltammetry (DPV)

Differential pulse voltammetry (DPV) involves varying the potential by applying a pulse method, in which successive potential steps (pulse) are applied to the working electrode, superimposing the pulse to a slow linear potential sweep, see Figure 6. In DPV, the baseline potential is held for a period of time before the application of the potential pulse, each potential pulse is fixed, and the current is recorded twice in each pulse period prior to the pulse and towards the end of the pulse. The selection of these two sampling points allows the decay of charging current to minimise its effect; thereby achieving high sensitivity.[101] Differences between current measurements at these two points is plot as a function of potential, see Figure 7. DPV decreases the detection limit compared to CV, and thus provides a more sensitive method for the electrochemical study of dilute and valuable samples.

Figure 6 Graph shows the potential change over time for DPV scan. The pulse is superimposed into the linear scan.

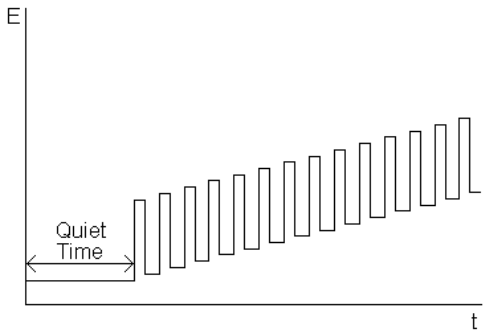
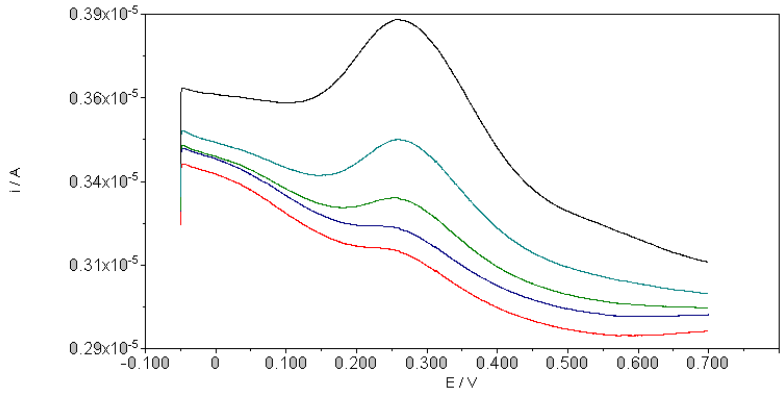


Figure 7 Typical DPV of 1,1'-s-ferrocenyldimethyl-L-glutathione at various concentrations between 11 to 84 µM. The linear sweep can only detect the ferrocene group.

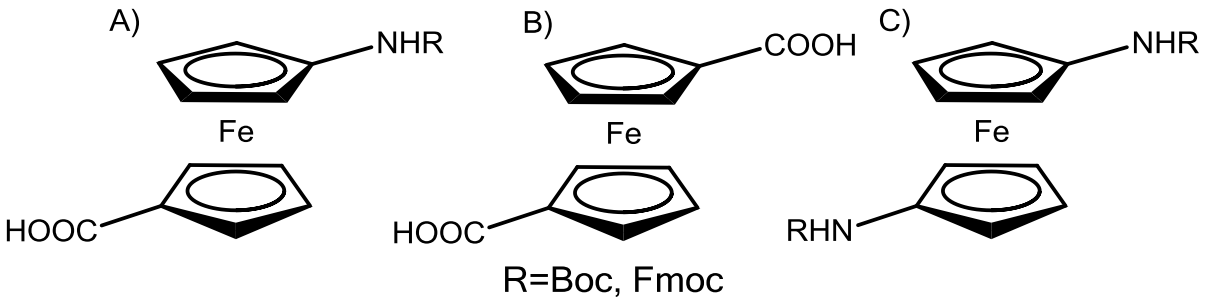


Chapter 3 Strategies for the dimerisation of peptide with ferrocene

3.1 Introduction

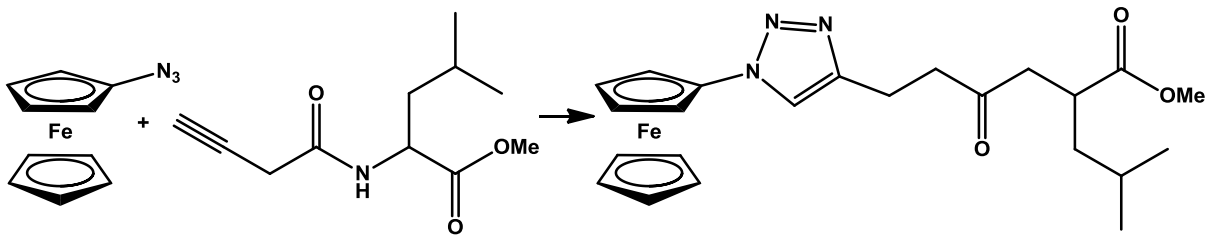
Due to ferrocene's aqueous stability and its characteristic electrochemistry, ferrocene has become a popular molecule for electrochemical sensing and for biological studies. The latter is generally achieved by coupling a ferrocene unit with a biomolecule of interest, which has been extensively studied.[102] Metzler-Nolte and Beckmann developed 1'-aminoferrocene-1-carboxylic acid (Figure 1) for use in solid-phase peptide synthesis (SPPS), and subsequently used the N-terminus protected ferrocene building block in both *tert*-butoxycarbonyl (Boc) and fluorenylmethoxycarbonyl (Fmoc) SPPS methods.[103,104] Peptides were coupled with 1'-aminoferrocene-1-carboxylic acid, 1,1'-diamino ferrocene and ferrocenedicarboxylic acid, in an effort to prepare peptide conjugates with antiparallel or parallel alignment of the peptide substituents.[103]

Figure 1 A) 1'-Aminoferrocene-1-carboxylic acid, B) ferrocene dicarboxylic acid and C) 1,1'-diamino ferrocene. Their Fmoc and Boc protected derivatives were used in SPPS to integrate ferrocene into A) a peptide backbone or B)/C) for parallel alignment of two peptide strands.



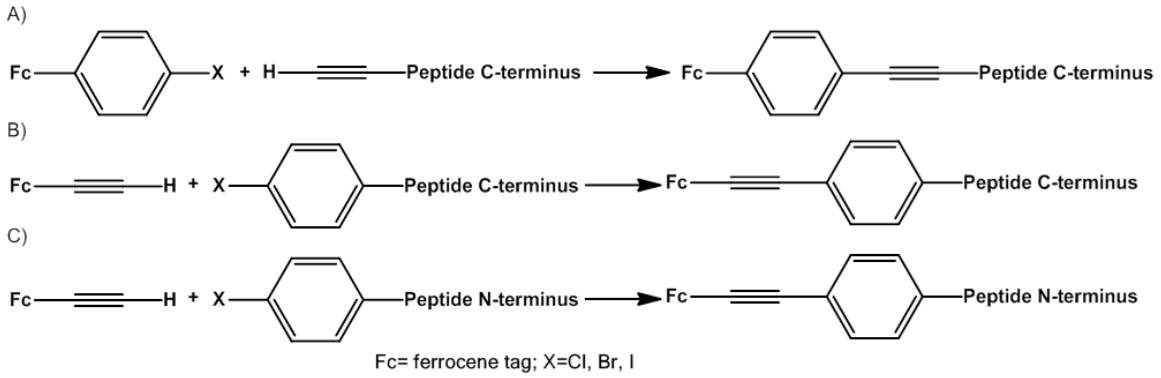
The 'click chemistry' method for ferrocene peptide preparation was recently reported (Figure 2), and is an attractive approach as synthesis is performed under mild aqueous conditions, and with high chemoselectivity.[55,105] However, one possible disadvantage involves the use on non-natural, and therefore more costly, amino acids.

Figure 2 'Click chemistry' method for coupling ferrocene azide with an alkyne.[55]



Sonogashira coupling is the cross coupling of an aryl halide and a terminal alkyne catalysed by Pd under an inert atmosphere, and it provides a method for the preparation of ferrocene peptide conjugates with low polarity covalent bonds.[55] Ferrocene can be tagged to either the C terminus or N terminus of peptides using this approach (see Figure 3).[106,107]

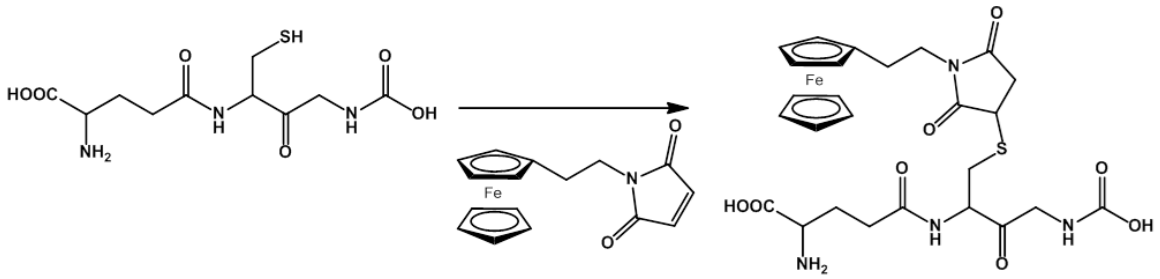
Figure 3 A) and B) Sonogashira coupling of ferrocene with the C-terminus of a peptide. The ferrocene unit can either be tagged to the aryl halide or the terminal alkyne for the cross coupling. C) Sonogashira coupling of ferrocene with the N-terminus of a peptide, only the ferrocene tagged terminal alkyne was reported.[55,106,107]



The labelling of ferrocenylethyl maleimide and ferrocene iodoacetamide with the thiol group of cysteine in glutathione and a longer peptide sequence, has been reported.[108-110] This approach provides a residue selective coupling strategy, without requiring any modification of the terminus of the peptide prepared by solid-phase peptide synthesis. Ferrocenyl N-substituted maleimide has been used to tag a hexapeptide in the literature.[111] The advantage of coupling through the thiol side

chain of cysteine is that this amino acid has low natural abundance in proteins and is often involved in disulfide bonds. This allows for the selective coupling to unique cysteine sites which can be engineered into peptides and proteins, which has the advantage of being a natural amino acid.

Figure 4 Coupling of the tripeptide glutathione with ferrocenylethyl maleimide through the thiol group on the cysteine. Both ferrocenylethyl maleimide and ferrocenylmethyl maleimide have been coupled with Cys in the literature.



3.2 Strategy design

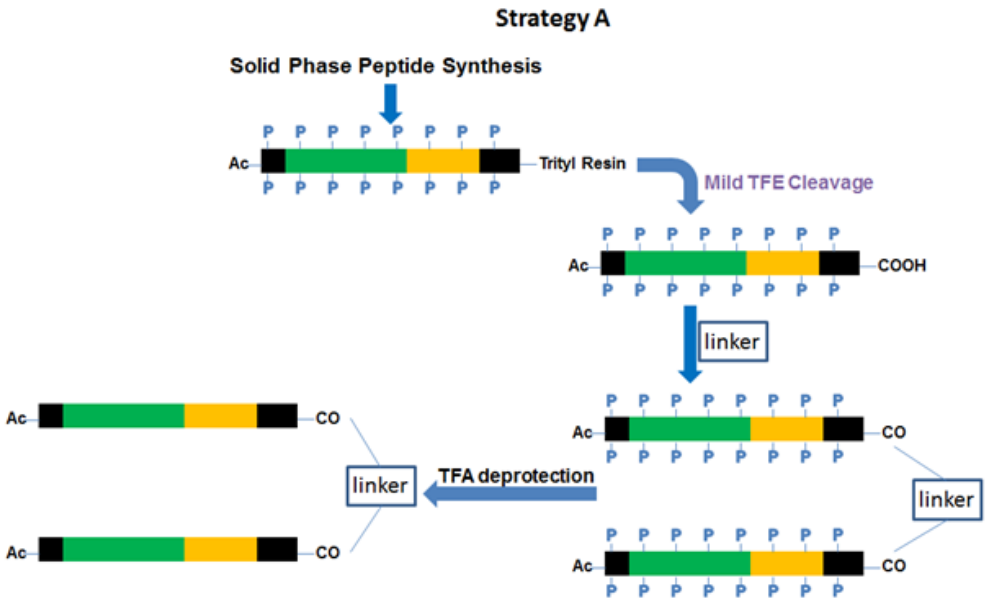
3.2.1 Strategy A

Two strategies were proposed to dimerise our peptides with a ferrocene unit. In strategy A, the peptide couples with the ferrocene linker through the free C-terminus (required due to the location of the leucine zipper at the C-terminus of GCN4). This therefore required the preparation of a peptide with all side chains having protecting groups in place, in order to prevent side reactions, see Figure 5.

In order to cleave a peptide from the solid support, yet retain the side chain protecting groups, 2-chlorotrityl chloride resin was used. This is because a mild cleavage cocktail, 2,2,2-trifluoroethanol (TFE), acetic acid and DCM (2:2:6), is sufficient to cleave the peptide from the resin, whilst leaving all side chain protecting groups on. The resulting peptide could be coupled with the ferrocene linker molecule to give the ferrocene protected peptide dimer. A final TFA deprotection step would be

used to cleave all the side chain protecting groups from the peptide dimer, to yield the desired ferrocene peptide dimer product (see Figure 5).

Figure 5 Synthetic strategy A; peptide cleaved from resin by a mild TFE mixture, and then coupled with the ferrocene linker through the C-terminus. This is followed by a TFA deprotection, which removes all the amino acid side chain protecting groups.



For strategy A, 1,1'-bis amino ferrocene and a 29 amino acid side chain protected peptide were synthesised, the latter manually in a microwave reactor. However cleavage of the peptide from the resin resulted in an insoluble peptide, and it was not possible to either purify or characterise it. In order to characterise the peptide, the insoluble protected peptide was subjected to strong cleavage conditions (95 % TFA, 5% thioanisole, 3% EDT and 2% anisole), in order to remove all side chain protecting groups. Analytical HPLC and ESI-MS of the deprotected peptide, confirmed the presence of a major product which was the proposed 29 amino acid sequence. In order to evaluate if this strategy was worth pursuing, we attempted to couple our fully protected 29 amino acid peptide with a stilbene linker containing two amine groups which was commercially available, see Figure 6. Unfortunately this was unsuccessful

and only the mono-substituted ferrocene peptide was observed by MALDI MS, and thus the synthesis of the 1,1'-bis(aminoethyl)ferrocene linker molecule for strategy A was abandoned at the reduction of 1, 1'-bis nitrilemethylene ferrocene to 1,1'-bis(aminoethyl)ferrocene, see Figure 7.

Figure 6 Coupling of 4,4'-diaminostilbene with side chain protected 29 amino acid peptide. The protecting group was then removed by TFA cocktail for characterisation.

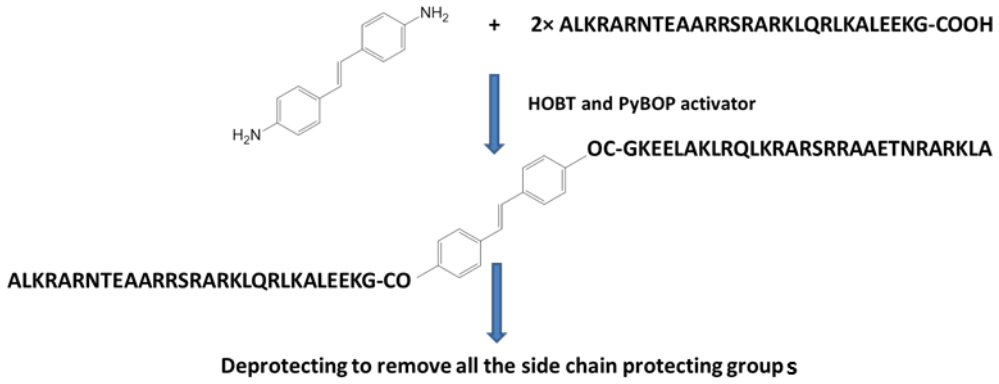
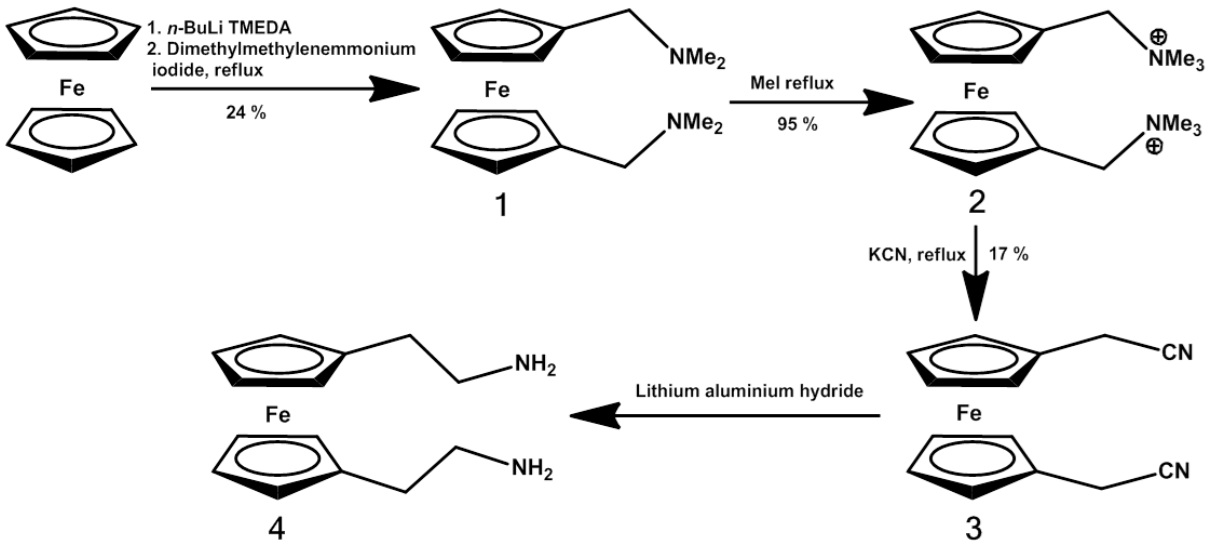


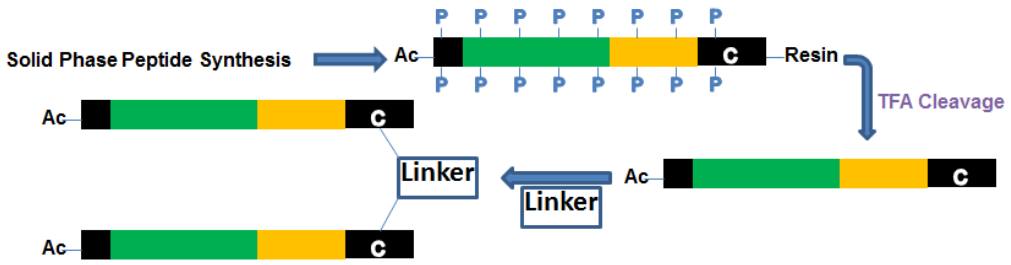
Figure 7 Scheme for the synthesis of 1,1'-bis(aminoethyl)ferrocene (4) from ferrocene in four steps. The synthesis was abandoned at the last step due to the failure of strategy A.



3.2.2 Strategy B

The requirement for a fully deprotected peptide (due to the insolubility of the protected analogue) meant that the peptide could be synthesised by solid-phase peptide synthesis using rink amide MBHA resin. Strategy B is illustrated in Figure 8. It therefore proved necessary to develop a new strategy for the dimerisation of two peptide domains. Due to the unique nature and low natural occurrence of the amino acid cysteine, we proposed to take advantage of the thiol side chain for selective coupling to our peptide. Strategy B therefore involves dimerising two fully deprotected peptides through a single cysteine introduced in each sequence (towards the C-terminus) for this purpose.

Figure 8 Strategy B: SPPS prepared peptide is cleaved and fully deprotected by TFA cocktail, and then the peptide coupled with a ferrocene linker molecule through its cysteine thiol side chain.



Initially we proposed to dimerise the peptide by coupling the cysteine side chain with a bromoacetic acid on 1,1'-bis bromoacetate methylene linker. 1,1'-bis bromoacetate methylene ferrocene was synthesised from ferrocene in four steps (see Figure 9), however it decomposed after column chromatography and HPLC. We therefore attempted to synthesise 1,1'-bis bromoacetate amino ferrocene, but only the mono substituted bromoacetate amino ferrocene was recovered in the product due to the rapid oxidation of the starting material 1,1'-bis-amino ferrocene to amino ferrocene in solution, see Figure 10.

Figure 9 Scheme for the synthesis of 1,1'-ferrocenyl bis (methylene 2- bromoacetate) (7) from ferrocene. The 1,1'-bis(hydroxymethylene)ferrocene (6) linker molecule was prepared in two steps and then reacted with bromoacetic acid to give the final linker molecule, 7.

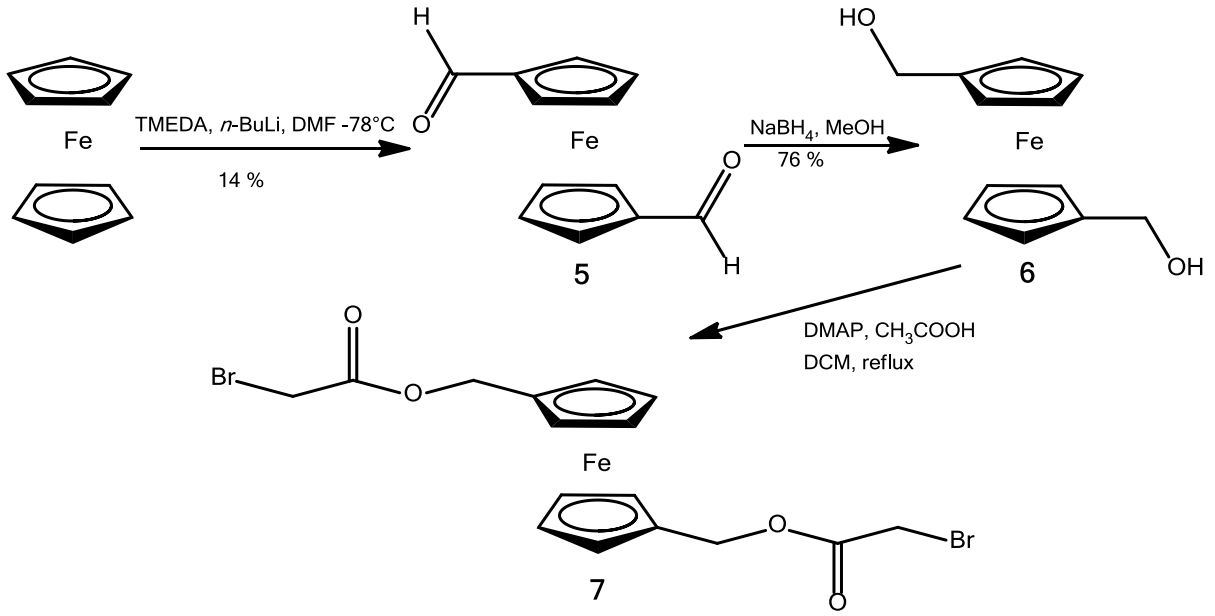
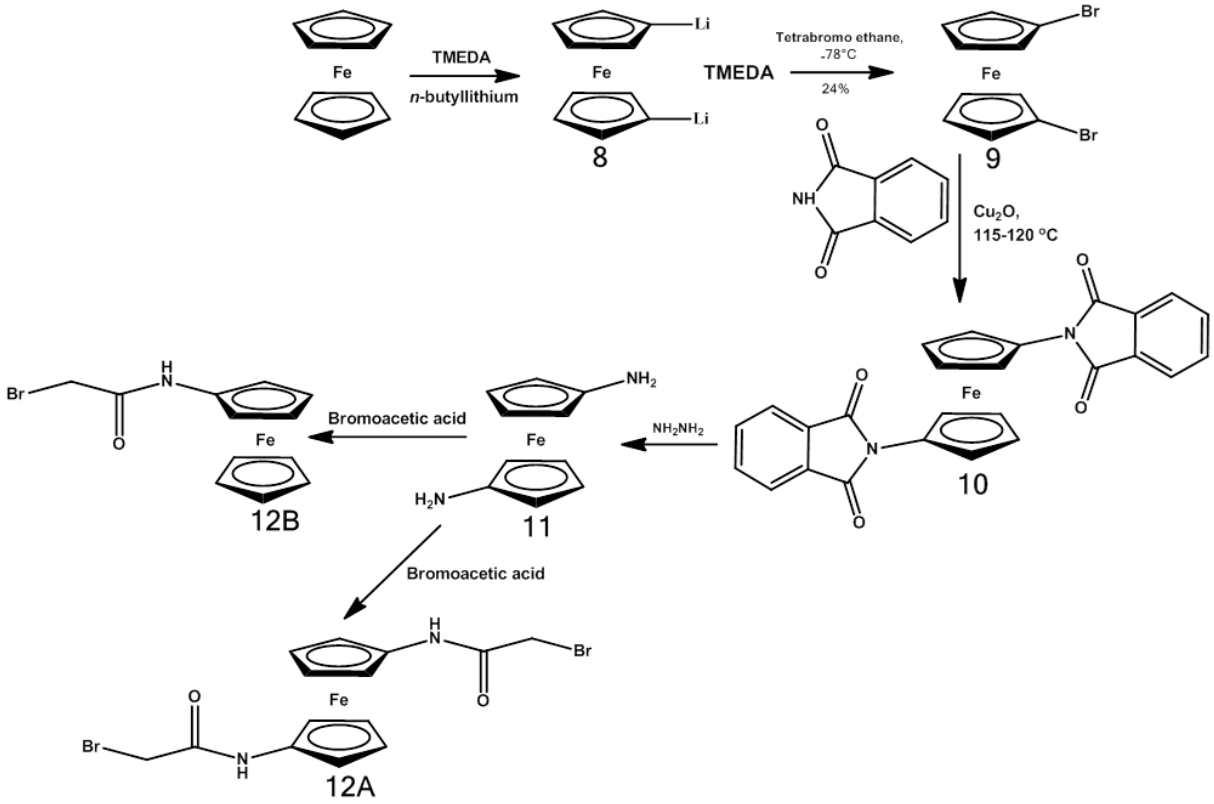


Figure 10 Scheme for the attempted synthesis of bis (2-bromo acetamide) ferrocene (12A) in five steps: lithiation of ferrocene, followed by bromination, two reactions for transformation into diamino ferrocene (Gabriel synthesis) and finally coupling with bromoacetic acid. However, only the mono substituted product 12B was formed.



The final method involves the coupling of a peptide cysteine side chain with 1,1'-bis(hydroxymethylene)ferrocene (6), synthesis in Figure 9, which can be synthesised from ferrocene in two steps. The ferrocene linker was first coupled with cysteine. In order to evaluate the reaction conditions required for efficient coupling, the resulting model complex was fully characterised in order to establish the influence of the Cys residue on the ferrocene chemistry. This work was subsequently expanded to longer peptides such as the Cys containing tripeptide glutathione (GSH). The resulting diGS model compound was compare to the diCys product, and represents a better model for ferrocene dimers of longer peptide sequences.

3.3 Results and discussion

3.3.1 Microwave assisted and full manual Solid-Phase Peptide Synthesis (SPPS) of protected peptide

In order to determine an effective method for the synthesis of a side chain protected peptide, 3 different synthesises were conducted to prepare a short 8 amino acid side chain protected peptide (Ac-LKACEEKG-COOH) on a 0.1 mmol scale, see table 1. The resulting products from the three methods were compared and the best method was subsequently used to prepare a 29 amino acid peptide based on GCN4.

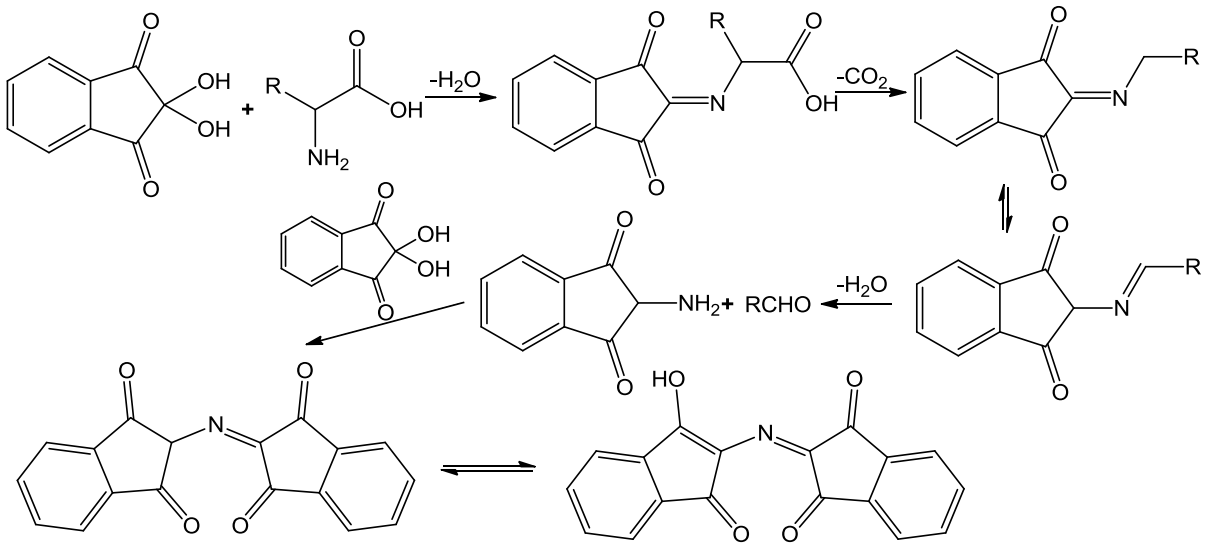
In method 1, the deprotection and coupling was assisted by microwave heating to 65℃; whereas in method 2, the deprotection and coupling was performed at room temperature and subsequently longer reaction times; in method 3, the deprotection and coupling was assisted by microwave heating to 50°C, but with longer reaction times than used in method 1 (Table 1).

Table 1 Synthesis of Ac-LKACEEKG-COOH on a 0.1 mmol scale. Methods 1 and 3 were carried out on a CEM microwave synthesiser. Method 2 was carried out manually at room temperature. HPLC purity was established using a C18 analytical column.

	deprotection	coupling	crude product yield	HPLC purity
method 1	65 ℃ 20 W 3 min	65 ℃ 20 W 15 min	17%	97%
method 2	RT 0 W 20 min	RT 0 W 120 min	23%	97%
method 3	50 ℃ 20 W 4.5 min	50 ℃ 20 W 18 min	25%	98%

For the synthesis of the protected peptide, 2-chlorotrityl chloride resin was chosen, as it can be cleaved from the resin using a mild TFE cocktail without removing the side chain protecting groups. The first amino acid Gly was loaded on to the resin manually. The loading rate of the first Gly was determined by monitoring the dibenzofulvene-piperidine adduct released on removal of the Fmoc group with piperidine, by UV visible spectroscopy based on a previously published procedure.[112] The loading rate of Fmoc-G-resin ranged from 0.65 mmol/g to 1.125 mmol/g in different tests for different samples. Better preparation of dry glass ware and the use of dry solvents, improved the reproducibility of loading rate measurements to 1.125 mmol/g. For the two methods of microwave assisted SPPS, a CEM microwave reactor was employed. For the manual SPPS, the synthesis was carried out in a glass reaction vessel equipped with a filter funnel, and each coupling and deprotection was monitored by the Kaiser Test.[113] In the commonly used Kaiser test,[113] ninhydrin reacts with free amines (the N-terminus of the peptide) to yield an intense blue colour ninhydrin imino derivative, see Figure 11.

Figure 11 The Kaiser test showing the reaction between the ninhydrin and the primary amine of an amino acid amine to produce a blue product.



3.3.1.A Comparison of three SPPS method for protected peptide preparation

Method 1 was carried out on a CEM discovery microwave synthesiser at 65°C, The yield is calculated from the fully deprotected peptide as a result of the TFA deprotection and cleavage; this was due to the lack of availability of TFE at this time. The crude product percentage yield is 17% and the H^+ and Na^+ adducts of the peptide was observed in the MALDI MS. Method 3 was also carried out on a CEM microwave synthesiser but at the lower temperature of 50°C.[114] Four different K^+ and Na^+ adducts of the protected peptide, cleaved by 20% TFE, 20% acetic acid and 60% DCM, were observed in the MALDI MS. The protected peptide was then fully deprotected by TFA, in order to establish a yield which can be compared to method 1, and to characterise it further. Method 2 involved manual synthesis at room temperature, each coupling and deprotection was monitored by the Kaiser Test with ninhydrin and phenol. In the protected peptide's MALDI MS generated by co-crystallising with the matrix gentisic acid, five different Na and K adducts of the

peptide were found. Full deprotection of the peptides prepared by methods 2 and 3 gave very similar percentage yields and high purities in HPLC (see table 1), however method 2 required longer reaction times with no additional benefit, and therefore method 3 was chosen for further synthesis of protected peptides.

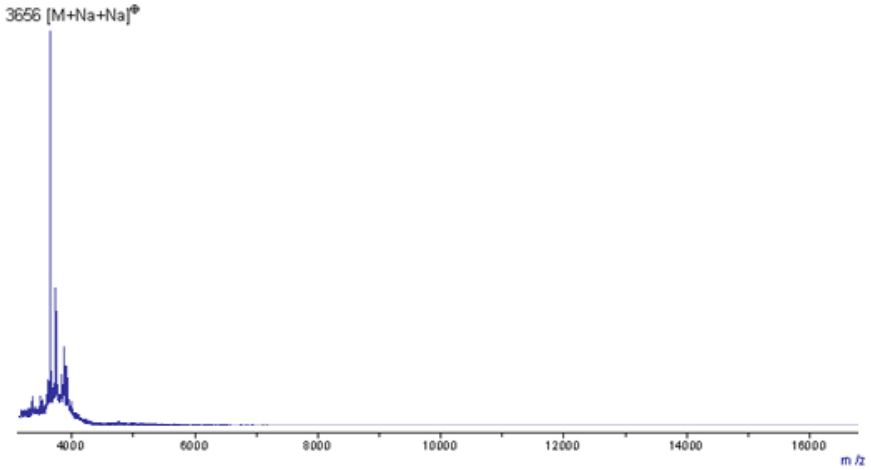
3.3.1.B Preparation of a 29 amino acid peptide based on GCN4

A 29 amino acid protected peptide sequence based on GCN4 with a C-terminus carboxylic acid (ALKRARNTEAARRSRARKLQRLKALEEKG-COOH) was designed to couple with a linker containing two amines. Microwave SPPS at 50 ℃ based on method 3 was applied after the manual loading of the first glycine (Gly G) residue. 1A, 3K, 4R, 5A, 6R, 7N and 8T were double coupled because they are located towards the end of the synthesis which makes the coupling difficult. All the arginines (Arg, R) were coupled at room temperature because microwave heating accelerates the formation of lactams during the carboxy-activation of arginine, which competes with the amide bond formation. 23K to 28K were therefore coupled at room temperature for 120 minutes. Before an aliquot was cleaved and analysed by C18 analytical HPLC and MS to validate the successful synthesis of the peptide sequence at this stage, and HOBt was not included in the activator. The activator mix HOBT and HBTU in DMF, was used instead of HBTU in DMF, after 23K. 21R was coupled by microwave at 50 ℃ by mistake and 10 mg of the resin was cleaved using the TFE cleavage mixture after coupling of residue 20Q. Its MALD MS indicated the Na and K adduct of protected Fmoc-QRLKALEEKG-COOH and therefore the correct peptide sequence. Another 10 mg of resin was cleaved after the coupling of 9E, and its MALDI MS indicates the H adduct of the protected peptide. The protected 29 amino

acid peptide has very poor solubility in organic solvents and water, and could not be detected by MALDI MS or purified by HPLC. Full deprotection by the TFA cleavage mixture resulted in a soluble peptide which could be analysed by MS. The H adduct of the peptide was found by MALDI MS, as well as a second peak at 1085 m/z which is also found in the MS after coupling 6E, and is consistent with the molecular weight of Ac-RLKALEEKG-COOH. The yield of crude 29 amino acid protected peptide is 30 %, which is higher than that achieved in the synthesis of the 8 amino acid peptide using method 3; the use of HOBT and HBTU activator could have improved the yield.

We attempted to couple the crude protected peptide with 4,4'-diaminostilbene dihydrochloride, which is commercially available, using HOBT and PyBOP activator. When a suspension of 4 equivalents of the crude peptide was used, only mono substituted stilbene peptide adducts were detected in the MALDI MS after TFA deprotection (Figure 12). This coupling reaction could not be monitored directly by MS or HPLC due to the poor solubility of the protected 29 residue peptide. When 8 equivalents of crude protected peptide was used for the coupling, neither the peptide dimer nor the mono substituted peptide was detected in MALDI MS following the TFA deprotection.

Figure 12. MALDI MS of the coupling reaction mixture (using 4 equivalents of crude protected peptide) after full deprotection using the TFA cleavage mixture. M is the mass of the mono-substituted stilbene peptide molecule. Except for the peptide monomer, no other peaks were observed in the lower range of the MS.



3.3.2 Synthesis of 1,1'-bis(cyanomethyl) ferrocene

We designed a synthesis of 1,1'-bis(aminomethyl) ferrocene from the reduction of the corresponding nitrile, which is prepared by the nitrile substitution of trimethyl ammonium, that is methylated from 1,1'-bis-dimethylaminomethyl ferrocene prepared from ferrocene (see Figure 7). Lithiation of ferrocene followed by electrophilic substitution of dimethyl methyleneammonium gave the bis substituted product in a 24 % yield, which is lower than that reported in the literature.[115] This is most likely due to a lack of a good nitrogen atmosphere and dry glassware preparation. *N*-alkylation of 1,1'-bis dimethylaminomethyl ferrocene by iodomethane gave high yields of 88-95 % of the bis ammonium ferrocene product. Nucleophilic substitution of the ferrocene ammonium by cyanide was carried out under various different conditions including under microwave irradiation and using 18-crown-6, see Table 2. However none of these resulted in a high yield.

Table 2 Different reaction conditions used for the substitution of ferrocene ammonium by cyanide performed either under reflux or in a CEM microwave reactor.

	Reagent	reaction condition	Yield
Synthesis 1	KCN	Heating under reflux in water for 2 hours	**none determined**
Synthesis 2	KCN	Heating under reflux in water for 6 days	**17 % isolated by solvent extraction**
Synthesis 3	KCN, 18-crown-6	Heating under reflux in MeCN for 6 days	**9 % isolated by solvent extraction**
Synthesis 4	KCN	Irradiation by microwave in water for 3 hours (100 ℃)	**none determined**

A nucleophilic substitution of the ferrocene ammonium by cyanide for the mono substituted ferrocene has been reported to be successful under reflux in water for 2 hours,[116] however this did not work for the bis substituted ferrocene, so a longer reaction time was attempted. This resulted in a low yield of the ferrocene nitrile product which was characterised by ES MS and ^{1}H NMR. In order to improve the yield, 18-crown-6 was used with KCN in acetonitrile as 18-crown-6 can complex the potassium in solution.[116] However it did not improve the yield.

Microwave irradiation could accelerate the reaction, but no nitrile product was detected. We observed our 1,1'-bis cyanomethyl ferrocene product darkened and decomposed to dimethyl ferrocene as indicated by ES MS, after being stored in the freezer at -20 ℃ overnight.

Due to the solubility problems encountered at this time with our 29 amino acid side chain protected peptide, which cannot be purified and characterised before TFA deprotection and thus its coupling with linker molecule cannot be monitored, the preparation of this linker was abandoned at this last step.

3.3.3 Dimerisation with ferrocene linker through the cysteine side chain

3.3.3.A Preparation of 1,1'-bis(hydroxymethylene)ferrocene and 1,1'-ferrocenyl bis (methylene 2- bromoacetate)

The ferrocene linker molecule **6** was prepared from ferrocene through 1,1'-ferrocenedicarboxaldehyde (**5**) in two reactions based on Connell's and Patwa's method, see Figure 9.[117,118] Recrystallisation gave **6** as brown crystals, which was later used as a linker for the coupling of peptides through the cysteine side chain (see below). 1,1'-Ferrocenyl bis (methylene 2- bromoacetate) (**7**) was synthesised by the esterification of 6 with three equivalents of bromoacetic acid catalysed by 4-dimethylaminopyridine. However, the product decomposed to 1,1'-bis methyl ferrocene and characterised by ^{1}H NMR spectrum and ES-MS after column chromatography and HPLC. The crude **7** which contains bromoacetic acid was characterised by ^{1}H NMR.

An attempt to couple crude **7** which contains bromoacetic acid, with a crude 28 amino acid peptide which contains a single Cys residue (13 equiv) in 50 % acetonitrile 50 % 100 mM phosphate buffer pH 7 solution for 20 hours, was unsuccessful as confirmed by the MALDI MS, in which only the 28 amino acid peptide starting material was found.

3.3.3.B Preparation of 1,1'-diamino ferrocene and 1,1'-bis (2-bromo-acetamide) ferrocene

1,1'-Diamino ferrocene (**11**) was designed to couple with side chain protected peptide, and was also used for the preparation of 1,1'-bis (2-bromo-acetamide) ferrocene, see Figure 10.

1,1'-Dibromo ferrocene (**9**) was synthesised by bromination of the crude 1,1'-dilithiaferrocene (**8**) intermediate prepared from ferrocene. **11** was synthesised from **9** by Gabriel synthesis. **11** was finally coupled with bromoacetic acid to make the linker molecule (see Figure 10).

The crude **9** was synthesised three times at 26 mmol, 69 mmol and 138 mmol scales, and the crude product was used for further synthesis directly as only a very small amount of ferrocene impurity was detected in the ^{1}H NMR.

10 was synthesised at 1.3 and 2.9 mmol scale with 3.26 and 6.5 equivalents of phthalimide. 3.26 equivalents of phthalimide only gave the mono-substituted product. 6.5 equivalents of phthalimide gave a mixture of mono and di-substituted product with the di-substituted product detected in the MALDI and ES-MS spectra. We attempted to purify the di-substituted product by column chromatography, but TLC indicated mono and di substituted products were in equilibrium. Synthesis on the 2.9 mmol scale gave the mono substituted product in a higher percentage yield (24 %) than the 1.3 mmol scale (10 %). The crude product was used directly for further synthesis. The synthesis of **11** was carried out twice with all the procedures performed under a nitrogen atmosphere under Schlenk conditions. Despite this, the yellow product immediately darkens in the solution after work up. ^{1}H NMR spectra were recorded while it was decomposing and amino ferrocene was detected. For the fresh yellow sample, **11** was found in the low voltage ES MS at 5 volt and by ^{1}H NMR spectroscopy, at higher voltages (10 and 15 volt) fragmentation occurs, and only mono substituted amino ferrocene and ferrocene were observed.

In the second synthesis, the crude product was used for coupling with three equivalents of bromoacetic acid immediately after the synthesis. Only the mono

substituted product, 2-bromo-ferrocenylacetamide, was observed in the ES-MS. This synthesis of this linker molecule was abandoned due to the instability of the 1,1'-diamino ferrocene.

3.3.3.C Synthesis of 1,1'-s-ferrocenyldimethyl-L-cysteine and 1,1'-s-ferrocenyldimethyl-L-glutathione

2.25 Equivalents of L-cysteine was coupled with **6** in water/acetone solution in the presence of 0.1 mL TFA. After 2 days of reaction the solvent was reduced in vacuo. The reaction progress was monitored by the Ellman's test to determine the concentration of unreacted thiol groups (from unreacted cysteine side chain), see Figure 13. The product was precipitated by addition of cold diethyl ether into a reduced volume of the reaction solution, to yield pure 1,1'-s-ferrocenyldimethyl-L-cysteine (**13**) as a yellow solid in a 23% yield. This was fully characterised by ^{1}H NMR, ^{13}C NMR, MALDI and ES mass spectrometry.

Figure 13 Concentration of free thiol in the cysteine coupling reaction with 1,1'-bis (hydroxymethylene) ferrocene. It is calculated from the Ellman's test of aliquots taken from the reaction solution at various time points. Plot showing the A) full reaction in 50 hours, and B) an expansion of the first 100 minutes of the reaction.

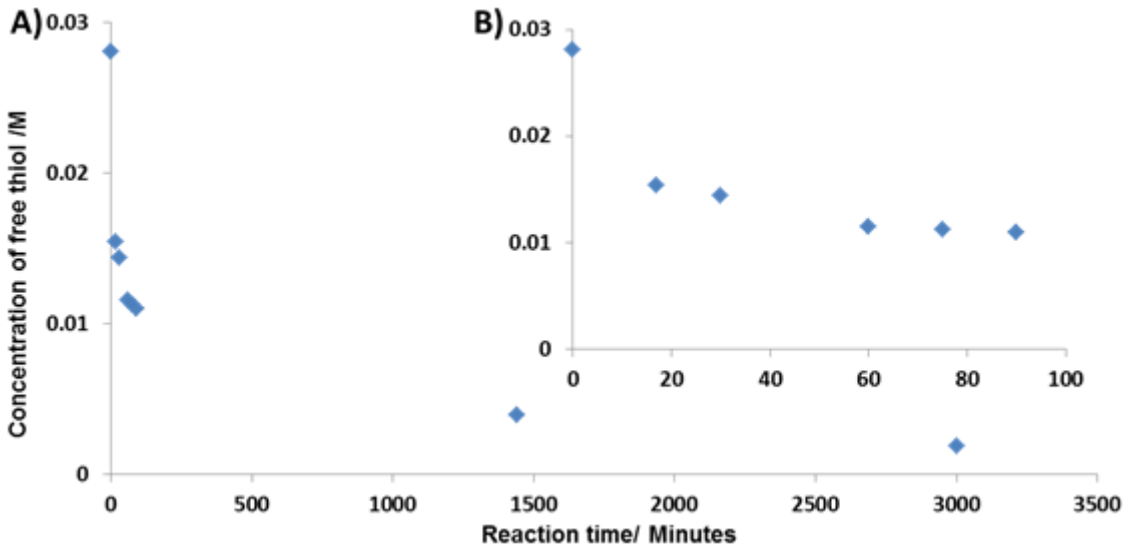


Figure 13 shows the change in cysteine concentration in the whole experiment up to 50 hours, as well as an expansion of the first 100 minutes. The concentration of available thiol, and thereby free cysteine, decreased to 0.002 M from 0.03 M over 50 hours. The reaction is quick at the beginning, in the first 17 minutes the concentration halved to ca. 0.014 M, and this is the reaction half-life. In the first 60 minutes, the cysteine concentration decreased 0.018 M. In the following 30 minutes, the concentration only decreased 0.007 M, and so the reaction has in effect reached equilibrium with no further significant change.

13 is not very soluble in either organic solvents or water. The addition of acid increases its solubility in water and so NMR spectra were recorded in the presence of acid, see Figures 14 and 15. The methylene hydrogens on the Cys side chain occur as AB systems in the ^{1}H NMR spectra with a geminal coupling constant J_{AB} of 15 Hz due to the chirality of the L-cysteine. After **13** has been in acidic solution overnight, the solution turn from yellow to green, and the peaks of the Cp ring hydrogens and the hydrogens adjacent to the Cp ring, broadened.

13 has also been characterised by PENDANT ^{13}C NMR spectrum, which determines the nature of primary, secondary, tertiary and quaternary carbon atoms, see Figure 15.

Figure 14 ^{1}H NMR spectrum of 1,1'-s-ferrocenyldimethyl-L-cysteine in MeOH-d_4 –HCl (95-5) (A) and DMSO-d_6-TFA (95-5) (B).

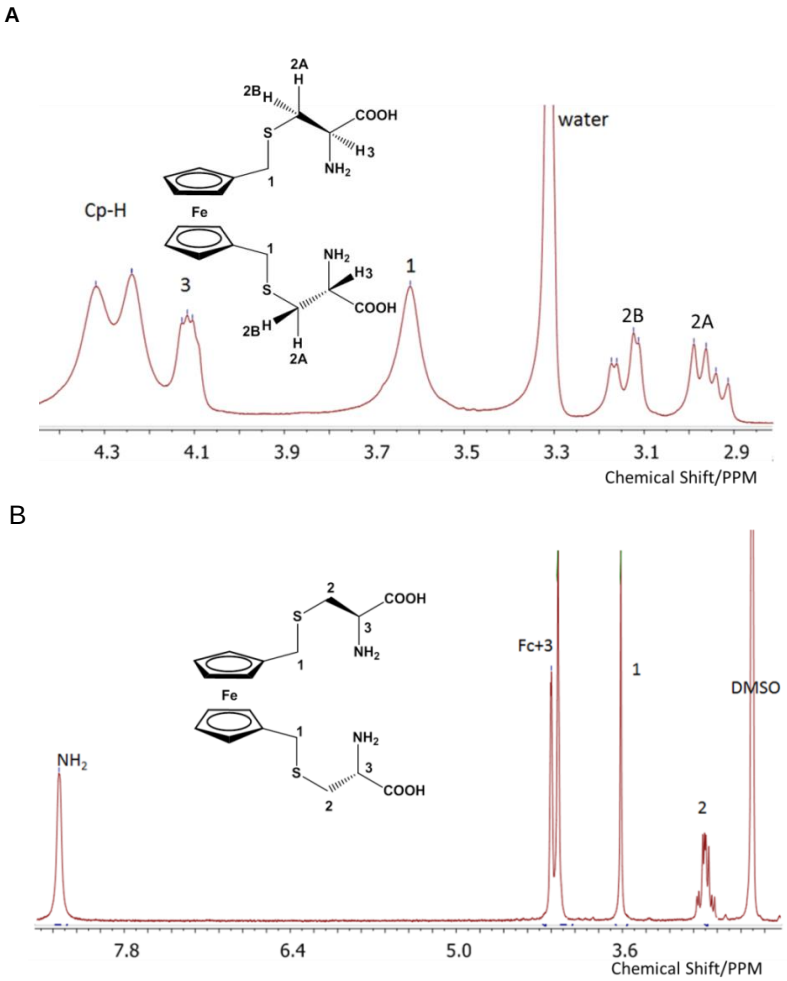
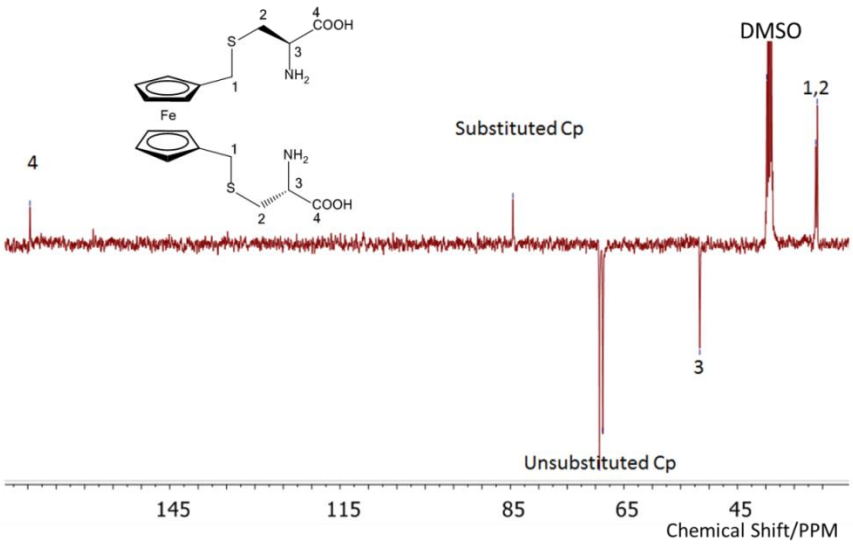


Figure 15 PENDANT ^{13}C NMR spectrum of 1,1'-s-ferrocenyldimethyl-L-cysteine in DMSO-d_4– TFA (95-5)



13 was also characterised by MALDI and ESI-MS. **13** and its K and Na adducts were observed in MALDI MS. H, K and Na adducts of **13** were observed in the ES MS, see Figures 16 and 17. Its isotopic distribution pattern is in good agreement with the calculated isotopic distribution for [$FeC_{16}H_{22}N_2O_4S_2$], see Figure 16.

Figure 16 MALDI MS of 1,1'-s-ferrocenyldimethyl-L-cysteine recorded in the presence of gentisic acid matrix. The peptide dimer and its Na and K adducts were observed. The experimental and calculated isotopic distributions are in good agreement.

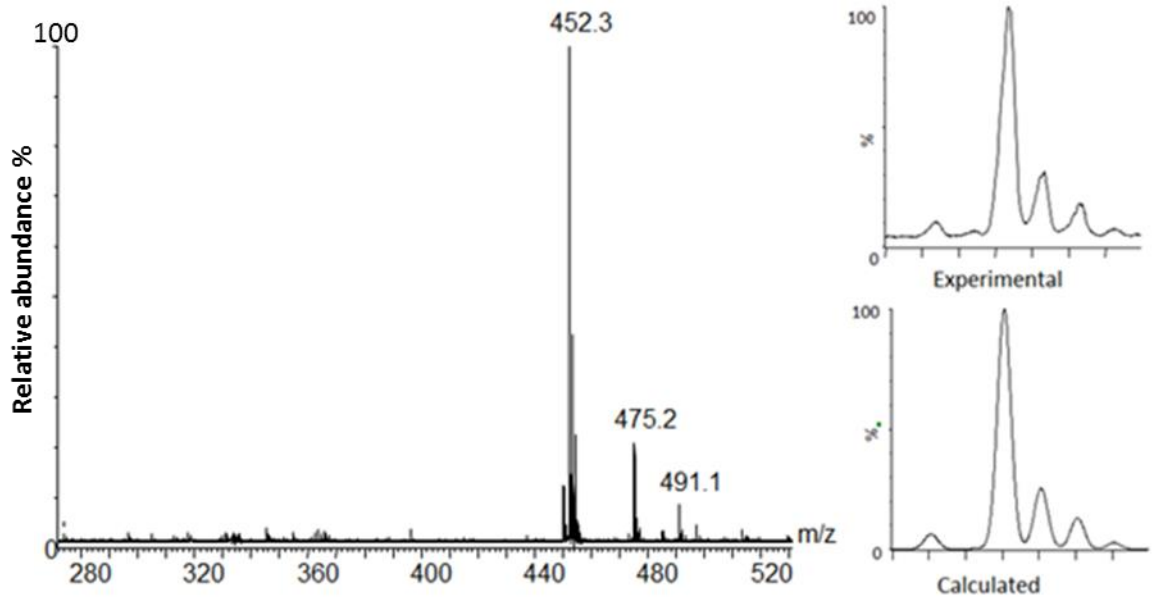
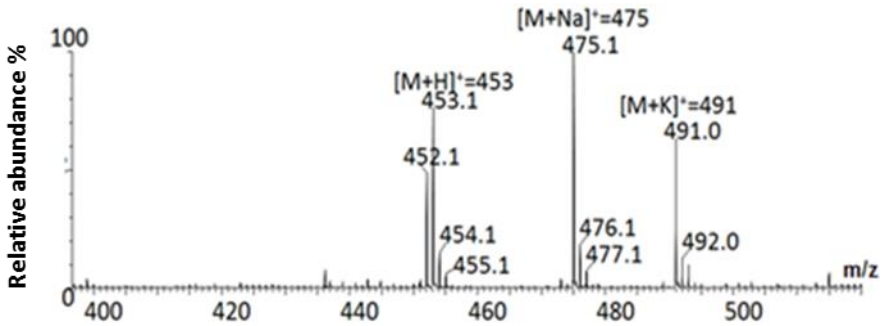


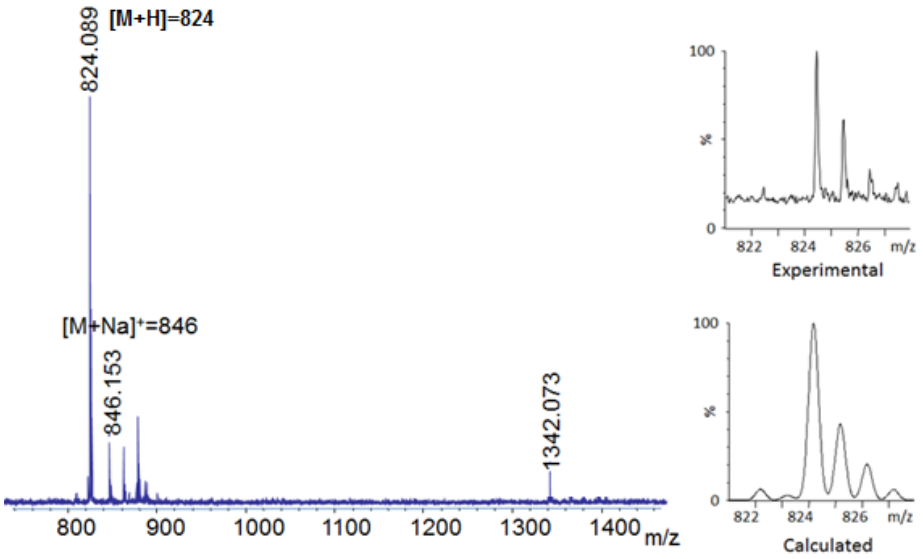
Figure 17 ES-MS of 1,1'-s-ferrocenyldimethyl-L-cysteine, recorded in positive ion mode. H, Na and K adducts of the product were observed.



L-Glutathione was coupled to **6** using the same method as described for the Cys coupling reaction. The product was purified by precipitation, by the addition of ethanol into an aqueous solution of the crude product, to yield an orange product.

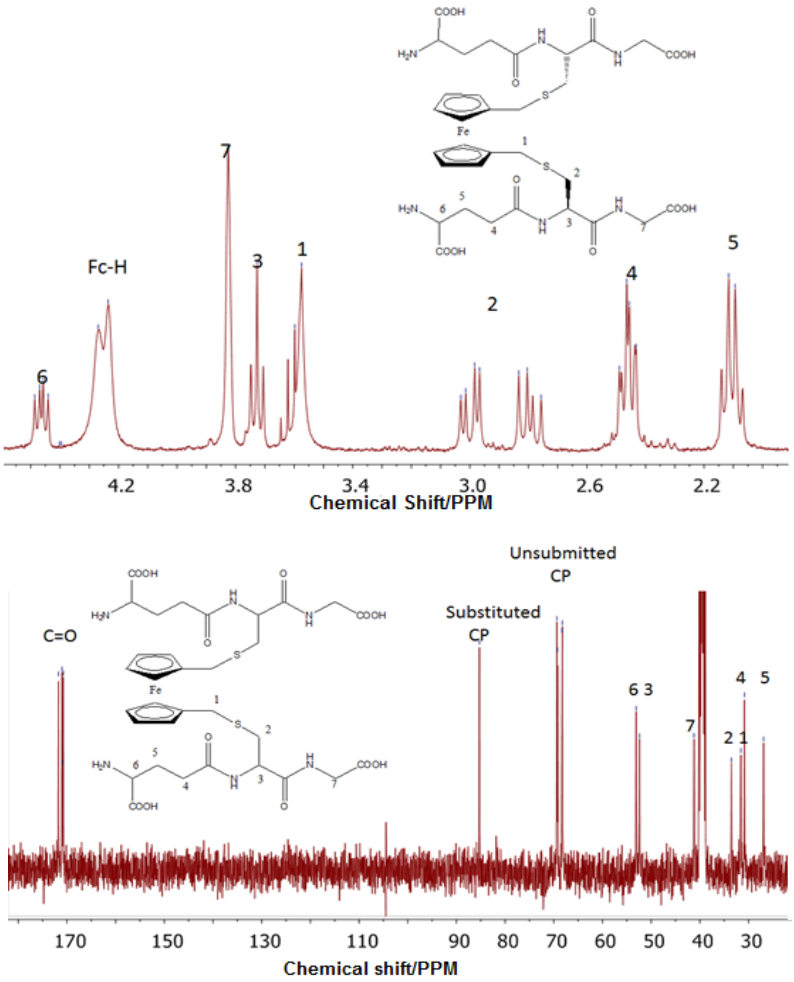
MALDI MS of 1,1'-s-ferrocenyldimethyl-L-glutathione (**14**) contained the desired product as the major species, as well as a small amount of a larger species which could be consistent with a disubstituted peptide coupled to a mono substituted peptide, see Figure 18. However, this species was not detected in the ^{1}H NMR and ^{13}C NMR spectra, and therefore could be an artefact formed in the gas-phase.

Figure 18 MALDI MS of 1, 1'-s-ferrocenyldimethyl-L-glutathione; hydrogen and Na adducts of the peptide were observed, 1342 is consistent with a disubstituted peptide coupled to a mono substituted peptide. The theoretical and experimental isotopic distribution patterns of the product are in good agreement.



1,1'-s-Ferrocenyldimethyl-L-glutathione contains 6 amino acids, including polar residues, and displays significantly better solubility that the smaller 1,1'-s-ferrocenyldimethyl-L-cysteine analogue. Its NMR spectra were all recorded without acid in the solution, see Figure 19. However it did precipitate out of water when stored at room temperature overnight. In the ^{1}H NMR of 1,1'-s-ferrocenyldimethyl-L-glutathione, again the methylene hydrogens on the Cys side chain occur as AB systems in the ^{1}H NMR spectra with a geminal coupling constant J_{AB} of 15 Hz due to the chirality of the L-cysteine., see Figure 19.

Figure 19 NMR spectra of 1,1'-s-ferrocenyldimethyl-L-glutathione. ^{1}H NMR was recorded in D_2O and ^{13}C NMR was recorded in DMSO-d_6. The peptide dimer has lower solubility in water.

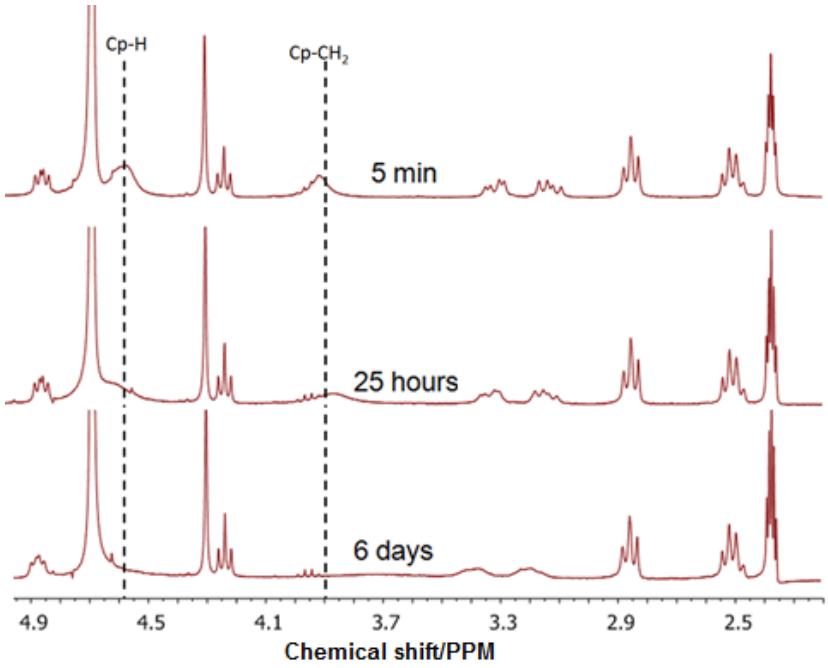


3.3.4 Stability test of 1,1'-s-ferrocenyldimethyl-L-glutathione by ^{1}H NMR spectroscopy

As the 1,1'-s-Ferrocenyldimethyl-L-glutathione precipitates out from water solution, and this water solution is required for many further experiments, acid was added to improve its longer term solubility. Diglutathione substituted ferrocene was dissolved in a water-TFA (95-5) solution, and this solution stored for 6 days without the appearance of any precipitation. A stability test in a MeCN-d_3:D_2O:TFA 50:50:0.05 solution, monitored by ^{1}H NMR, indicates that the Fe^{2+} to Fe^{3+} oxidation process is occurring in the ferrocene over these 6 days, as the peaks assigned to the

hydrogens on the Cp ring and those next to the Cp ring, become broader (see Figure 20), which imply the formation of the paramagnetic ferrocenium. Acetonitrile was used in the solution mixture for the stability test because this solvent is adopted in HPLC purification of peptides, and in addition the acetonitrile can provide a peak for calibration of ^{1}H NMR spectra.

Figure 20 ^{1}H NMR spectra of 1,1'-s-ferrocenyldimethyl-L-glutathione in a MeCN-d_3:D_2O:TFA 50:50:0.05 solution, recorded over 6 days. Peaks assigned to the Cp hydrogens and hydrogens next to Cp ring, broaden over time.

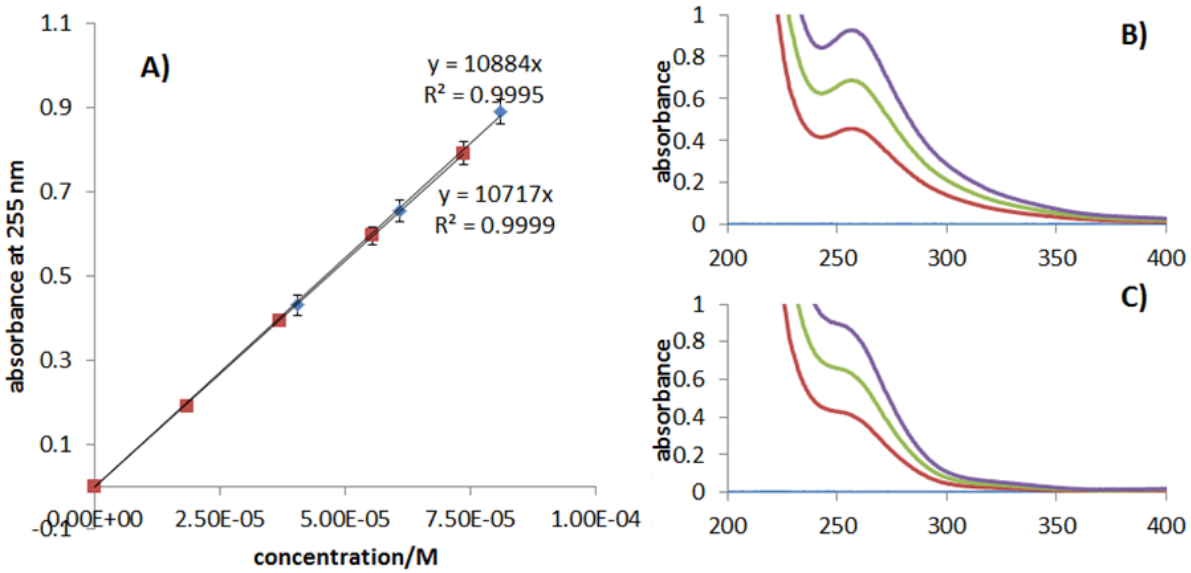


3.3.5 UV of 1,1'-s-ferrocenyldimethyl-L-glutathione and 1,1'-s-ferrocenyldimethyl-L-cysteine

Ferrocene peptide dimers have a peak centered around 255 nm in their UV-visible spectra (Figure 21). UV-visible spectra of the dimers at various concentrations allow us to determine the extinction coefficient of ferrocene which is used in subsequent chapters for concentration determination of the longer ferrocene peptide dimers. Therefore, solutions of both **13** and **14** were prepared at different concentrations and

their UV-visible spectra recorded. Experiments were repeated in triplicate in order to validate that data and their average absorbance values at 255 nm were plot as a function of concentration, from the slope of which their extinction coefficients could be determined (see Figure 21).

Figure 21 A) The average absorbance of the dimers at 255 nm is plotted against the concentration, the error bars correspond to standard deviations generated from three repeated experiments. UV-vis spectra of C) 1,1'-s-ferrocenyldimethyl-L-glutathione and B) 1,1'-s-ferrocenyldimethyl-L-cysteine in water-TFA (97-3) at three different concentrations at room temperature.

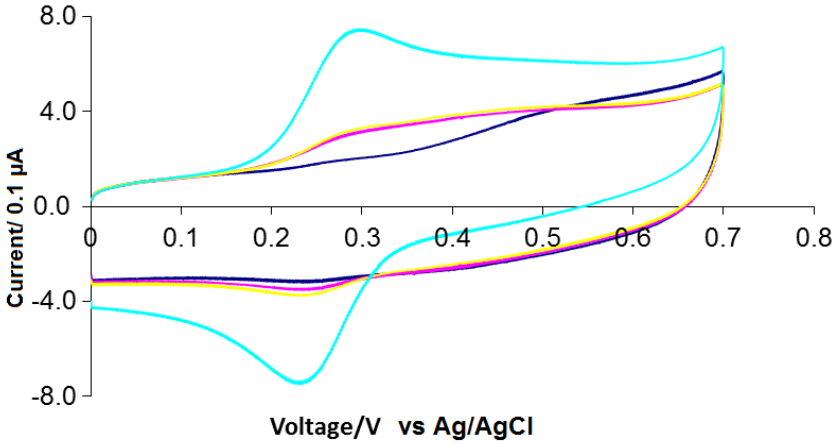


Both dimers have similar extinction coefficients of 10884 and 10717 M^{-1} cm^{-1} as expected. The average value was therefore used for experiments with the longer ferrocene peptide dimers, see Chapters 4 and 5.

3.3.6 CV titration of 1,1'-bis(hydroxymethylene)ferrocene

1,1'-Bis(hydroxymethylene)ferrocene (**6**) stock solution was titrated into a 100 mM phosphate buffer pH 7 solution, and the CV recorded between 0 and 0.7 V using gold and carbon electrodes, see Figure 22.

Figure 22 CV of 1,1'-bis(hydroxymethylene)ferrocene at 10, 20, 30 and 40 μM using a gold working electrode.



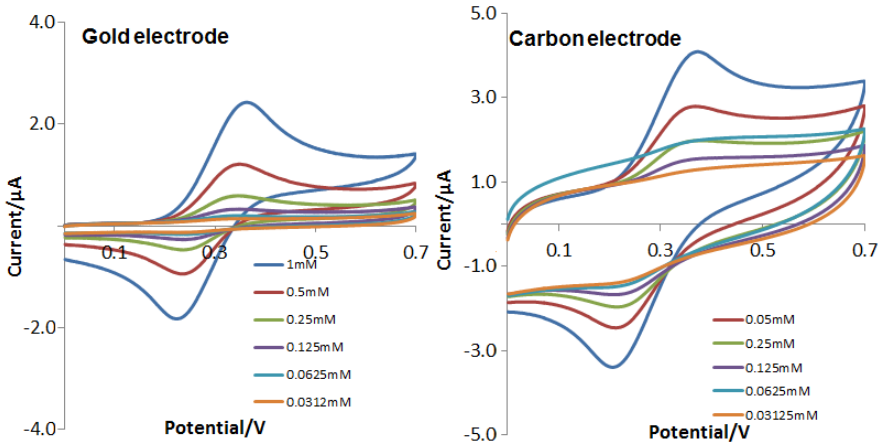
The ferrocene peptide dimer solutions at concentrations lower than 40 μM do not display proportional i_{pa} or i_{pc} to their concentration, because they are lower than the minimum detectable concentration of the experimental kit. 40 μM **6** solution gave a strong signal in the CV, the oxidation and reduction peaks of which are easily defined, and i_{pa}/i_{pc} (peak anodic current and peak cathodic current) is equal to 1, so **6** is considered to be chemically reversible. The ΔE_p (difference between oxidation potential and reduction potential) is equal to 57 mV, which indicates a single electron redox process, as an ideal one electron process should yield a ΔE_p of 59 mV if it is electrochemically reversible.[119]

3.3.7 CV of 1,1'-s-ferrocenyldimethyl-L-glutathione (diGSCH$_2$Fc) solutions

A titration of 1,1'-s-ferrocenyldimethyl-L-glutathione (**14**) was used to check the experimental kit with a low-volume sample holder, in which the counter and reference electrodes are not in the sample holder, but in the bulk electrolyte solution, coupled to the sample through a glass frit (see Chapter 2 for full experimental details). The

experiment was conducted using a gold and carbon working electrode, respectively, see Figure 23.

Figure 23 CV titration of 1,1'-s-ferrocenyldimethyl-L-glutathione using a carbon and gold working electrode, and a small volume fritted sample holder.

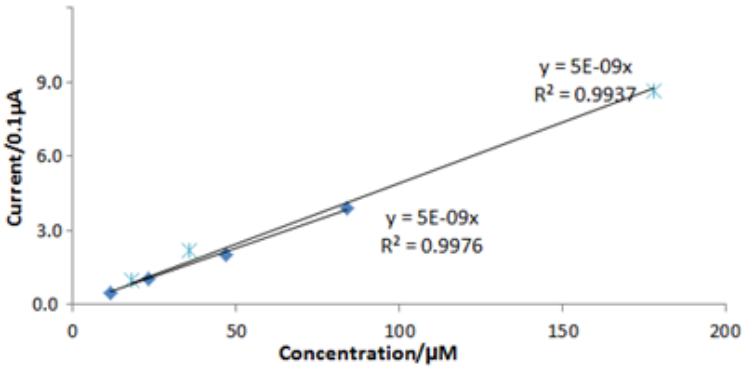


In both the CV recorded with either a gold or carbon working electrode, the i_{pa} and i_{pc} are both proportional to the concentration. However, the i_{pa} is always larger than i_{pc}, which indicates it is not chemically reversible in this experimental kit. The ΔE_p is around 100 mV, which is much larger than the ideal value of 59 mV, which indicates it is not electrochemically reversible. Alternatively this could be because there is a problem with the conductivity at the frit of the sample holder where the electron transfer happens between the sample solution in the holder and the buffer solution in the larger vial. We continued the comparison of these two electrochemistry experimental kit set-ups using our longer ferrocene peptide dimers, and these findings are reported in Chapter 4.

3.3.8 DPV titration of 1,1'-S-ferrocenyldimethyl-L-glutathione

A DPV titration was performed in addition to the CV experiments, as DPV is more sensitive and can therefore be recorded on solutions of lower concentrations (important when working with biomolecules). The titration was performed twice at various concentrations, and the resulting peak currents plotted as a function of concentration (see Figure 24). Both DPV titrations of 1,1'-S-ferrocenyldimethyl-L-glutathione share the same gradient. When the concentration is 5.67 µM and 7.14 mM, it did not give valid peak currents on the plot, but the plot did suggested that a concentration range of between 11.5 to 178 µM should yield valid data in the DPV experiment.

Figure 24 Peak currents from two DPV titrations are plotted as a function of the concentration of 1,1'-S-ferrocenyldimethyl-L-glutathione. Both plots share very similar gradients with R^2 greater than 0.99.



3.4 Conclusions

We have developed an efficient method for the SPPS of side chain protected peptides by the synthesis of a short peptide using three methods at room temperature or with microwave irradiation. A side chain protected 29 amino acid peptide design based on GCN4 was prepared by this method for strategy A; it has low solubility in water and organic solvents and cannot be purified and characterised

without full deprotection of amino acid side chains. We attempted to couple this peptide with a commercially available stilbene linker, but only the mono substituted peptide was found in MS after deprotection. Due to the low solubility of peptide with fully protected side chains, strategy A was abandoned. However an unstable linker, 1,1'-diamino ferrocene, was synthesised for strategy A. The synthesis of 1,1'-bis aminoethyl ferrocene was abandoned at the last step, due to the required change in strategy. Strategy B was to couple a ferrocene linker with a single cysteine introduced into a peptide sequence for this purpose. 1,1'-Ferrocenyl bis (methylene 2-bromoacetate) was synthesised for strategy B, but it decomposed to 1,1'-ferrocenyl bis methylium after HPLC and column chromatography. We attempted to couple the crude 1,1'-ferrocenylbis(methylene-2-bromoacetate) with peptide, but no peptide dimer was observed in MS. We attempted to couple diamino ferrocene with bromoacetic acid, but only mono substituted product was found in ES MS due to the decomposition of diamino ferrocene immediately in solution. 1,1'-Bis(hydroxymethylene)ferrocene was coupled with cysteine and glutathione and the products have been fully characterised. Both the cysteine and glutathione coupled product have an absorbance at 255 nm in their UV-visible spectra, from which an extinction coefficient was determined. CV experiments of 1,1'-s-ferrocenyldimethyl-L-glutathione and 1,1'-bis (hydroxymethylene)ferrocene were performed using two different experimental set-ups in order to check the use of a small volume sample holder. 1,1'-Bis(hydroxymethylene)ferrocene was found to be electrochemically reversible in CV by the kit set-up with all three electrodes emerged in the sample solution. The two electrochemistry experimental set-up will be further compared for the longer ferrocene peptide dimers, described in Chapter 4.

Chapter 4 Design, synthesis, physiochemical characterisation and electrochemical evaluation of ferrocene peptide dimers based on GCN4

4.1 Introduction

GCN4 binds specific DNA target sites, as a helical homodimer with the binding affinity being highly dependent on dimerisation. GCN4 is dimerised through a C-terminal leucine zipper domain. However, the monomeric GCN4 basic region (which lacks a leucine zipper domain) can also recognise and bind to the target DNA site, but with a lower affinity.[120] Stabilisation of the GCN4 helical content and enhanced dimerisation, significantly improves the basic domain's affinity for the DNA target site. Which is why monomeric GCN4 has a much weaker DNA binding affinity than the bZIP domain of natural GCN4.[121,122]

An important study by Kim and co-workers involved replacing the leucine zipper domain with a disulfide bridge. They found that this dimer displayed similar DNA binding to those with the leucine zipper, whereas the monomer displays only weak DNA binding. This work therefore demonstrated that dimerisation was important for strong and selective DNA binding, and the role of leucine zipper was to function as a dimerisation domain.[35]

As dimerisation is the only function of the leucine zipper component of GCN4, peptides designed based on GCN4 and which have been coupled with non-natural linkers, have been reported in the literature.[35,36,38] These peptide dimers display similar DNA binding as native GCN4.

Mascarenas and co-worker synthesised a DNA binding peptide dimer based on basic region of GCN4 with a light-modulated azobenzene linker.[36] In their design, the peptide is dimerised with the azobenzene linker molecule through the Cys side chain of a Gly-Gly-Cys sequence, introduced at the C-terminus of the basic domain of GCN4. The azobenzene adopts a trans conformation, however an irradiation of 365

nm wavelength light, this converts to the cis conformation. The cis isomer binds to DNA containing the CREB recognition site, with 60-70 times more efficiency than the trans isomer, due to the more appropriate distance between the two DNA binding basic regions.

Schepartz and co-worker prepared a series of DNA binding peptide dimers which are composed of two identical 29 amino acid peptides contain the basic region and hinge of GCN4 and a Gly-Gly-Cys C-terminal cap linked by disulfide bonds to 2 different positions of a bis(terpyridine) iron (II) complex. The resulting complexes position the two DNA binding basic regions at different distances. These dimers were shown to display different binding affinities for the DNA containing GCN4 target site.[38]

The electrochemistry of ferrocene has been well studied, and a number of excellent reviews have recently been reported on it.[38,123,124] It has been used to tag large biomolecule for sensor purposes. The two Cp (cyclopentadienyl) rings of ferrocene can freely rotate due to the low energy barrier, which could provide important flexibility to allow peptide domains to be appropriately aligned for strong DNA binding. Ferrocene is stable to high temperature and unaffected by water.[125] From the crystal structure of ferrocene, the distance between the two Cp rings is 3.3 Å,[126] which is shorter than the distance between the side chains on the two Cys which replace the first residue in the leucine zipper domain in the crystal structure of GCN4-CRE DNA complex (Figure 1), and thus a flexible new spacer is required between the ferrocene and the peptides. 28AA peptide (Table 1) was therefore prepared which contained the full basic domain and hinge region of GCN4 terminating with a flexible Gly-Gly-Cys-Gly towards the C-terminus. A shorter 26AA peptide was also prepared which

differed from the 28AA peptide in that it lacked the two flexible Gly residues between the hinge and Cys for subsequent dimerisation with ferrocene. A 25AA peptide was prepared with one fewer Gly at the C-terminus, adjacent to the Cys residue, than the 26AA peptide. This offered the opportunity to study how the C-terminal Gly affects the dimer's DNA binding. Finally shorter 22AA and 24AA peptides were prepared, each of which contained fewer residues from the hinge region of GCN4 (LQRMKQ). The C-terminal Gly between the hinge region and the Cys residue, was introduced to make DNA alignment of the folded basic regions (α-helical recognition units) more flexible in the presence of target DNA. The 22AA and 24AA were compared so as to establish the shortest hinge sequence required to maintain the spatial relationship between the basic domain and the dimerisation unit, for effective DNA binding.

Figure 1 Distance measured between the side chains of Cys residues which replace the first residue in leucine zipper domain of GCN4-DNA complex crystal structure (PDB code: 2DGC).[29]

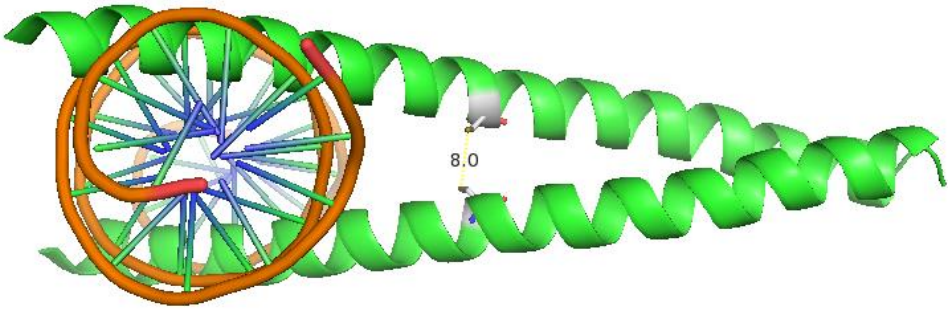


Table 1 Peptide sequences designed as DNA binding dimers based on GCN4.

	N-terminus	basic domain	hinge	C-terminus
wild type GCN4	MIVPESSDPAAL	KRARNTEAARRSRARK	LQRMKQ	EDKV[...]ER
Disulfide bridge[4]	PESSDPAAL	KRARNTEAARRSRARK	LQRMKQ	GGC
azobenzene linker[5]	DPAAL	KRARNTEAARRSRARK	LQ	GGC
Iron complex linker[6]	SAAL	KRARNTEAARRSRARK	LQRMKQ	GGC
MBR-CC DEGRADO[11]	AL	RRARNREAAARSRARR	AEKLKA	LEEK[...]LG
22AA	AL	KRARNTEAARRSRARK	L	GCG
24AA	AL	KRARNTEAARRSRARK	LQR	GCG
25AA	AL	KRARNTEAARRSRARK	LQRMKQ	C
26AA	AL	KRARNTEAARRSRARK	LQRMKQ	CG
28AA	AL	KRARNTEAARRSRARK	LQRMKQ	GGCG

4.2 Peptide sequence design based on GCN4

We compared peptide sequences used in the artificial miniature DNA binding peptides based on GCN4 which has previous been reported in the literature. [29,35,36,38] DeGrado and co-workers reported a MBR-CC, in which an artificial basic region only contained those residues which were conserved in the basic regions of eight different DNA binding bZIP proteins which bind to either the TRE or CRE sites in DNA.[29] In MBR-CC, only two residues on the C-terminus were retained to cap the basic region; and this has also been applied in our sequence design.

In the azobenzene linker GCN4 dimer, Mascarenas and co-workers carried out modeling by reference to Richmond's GCN4-CRE complex crystal structures to predict the right position to insert azobenzene linker so as for the cis and trans isomers to bind differently to DNA CRE site.[36,126]

4.3 Results and discussions

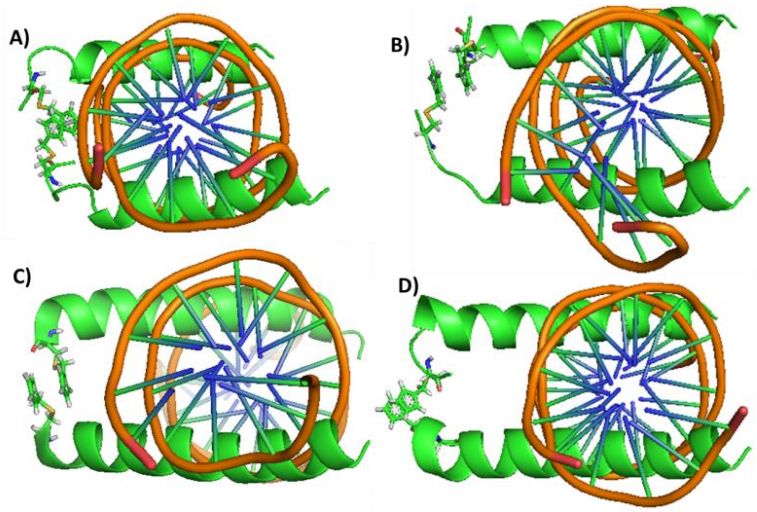
4.3.1 Molecular dynamic (MD) simulations of the peptide dimers

The MD simulations were performed based on the modified crystal structure of complex including GCN4 and CRE DNA.[126] The excess residues on the GCN4 peptide model were removed, N-terminal groups and C-terminal Cp ring tagged cysteine were put into the correspond position on the peptide. Parameter of the ferrocene-cysteine unit was imported to define the ferrocene group structure. The solvent box TIP3PBOX 15 and Na^+ were employed to the center of peptide-DNA complex to model the water solvent. The simulations were all performed with an energy minimisation followed by MD simulation. All the simulation stages were carried out with the all atom force field. In the initial 0.5 ns MD simulation, the

simulation temperature was set up to 298 K from 0K and then continued for the rest of the simulations up to 10 ns. Each complex simulation was performed in triplicate to validate the result and the average model was calculated as a final result. The 44AA, 48AA, 52AA and 56AA complex with CRE DNA were used for MD simulations.

In the MD simulation model of the 44AA-CRE complex after 10 ns, the C-terminus of the two peptide strands were unfolded up to the Lys residue at position 18, the only residue in the hinge domain, and the remainder of the peptide sequence (the basic domain) was α-helical and bound to the major groove of the DNA target site (Figure 2A). The single amino acid in the hinge sequence is too short for the dimeric DNA binding of peptide. In the MD simulation model of the 48AA-CRE complex after 10 ns, the C-terminus of one peptide strand is unfolded up to the Arg residue at position 21, three residues into the basic region (Figure 2B). 48AA forms a better complex with CRE DNA compared to 44AA, due to the longer hinge sequence. In the 52AA-CRE DNA complex after 10 ns, both strands are fully α-helical resulting in a well folded peptide-DNA complex. However, one peptide strand appears to have been pushed away from the DNA backbone in the 56AA-CRE DNA complex. Therefore I conclude that the sequence in 52AA can provide the optimal conformation for formation of a well folded peptide-DNA complex while the sequence in 56AA is too long for it to be reasonably accommodated, see Figure 2.

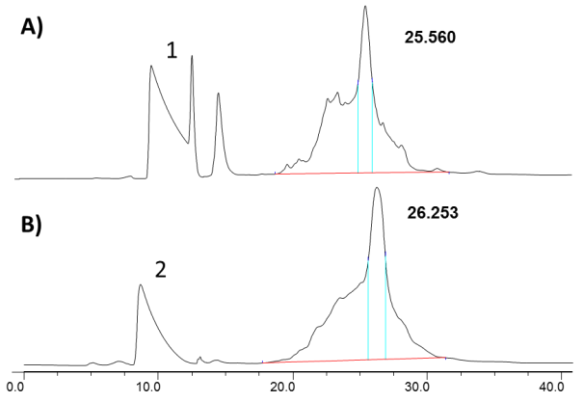
Figure 2 MD simulation models of peptide-CRE DNA complexes at 10 ns; A) 44AA, B) 48AA, C) 52AA and D) 56AA.



4.3.2 Method development for peptide synthesis on the Liberty-1 microwave automated peptide synthesiser

We attempted to synthesise the 26AA using a single amino acid coupling method as well as a method which included a single amino acid coupling followed by capping for each amino acid, however, both of these methods resulted in a very impure mixture in which the target peptide was not the major species. Instead various shorter peptide sequences were detected due to incomplete peptide synthesis. Therefore, we adopted the double coupling method for each amino acid of 26AA and the longer 28AA which yielded the target peptide as the major product, which was 28-32% pure as determined by preparative HPLC after cleavage from the resin, see Figure 3.

Figure 3 Preparative C12 HPLC of the crude A) 26AA and B) 28AA peptides, using a gradient increasing 0% acetonitrile (0.05 %TFA) in water (0.05 %TFA) to 20% over 20 minutes followed by a 20 minute wash of 20 % acetonitrile (0.05 %TFA) monitored at 222nm.



The initial attempt to cleave the 26AA from the resin using the TFA cleavage cocktail for a period of 2 hours led to the presence of an impurity. This was confirmed by MALDI MS (Figures 4 and 5) to be the 26AA with a Cys which still contained the triphenylmethyl protecting group. The peptide was then subjected to fresh cleavage TFA cocktail for a further 2 hours. This resulted in fully deprotected 26AA as the major product as confirmed by both HPLC and MALDI MS (Figure 4 and 5).

Figure 4 Analytical C12 HPLC of the 26AA after 2 hours and 4 hours cleavage respectively using a gradient increasing 0% acetonitrile (0.05 %TFA) in water (0.05 %TFA) to 20% over 20 minutes followed by a 20 minute wash of 20 % acetonitrile (0.05 %TFA) monitored at 222nm.

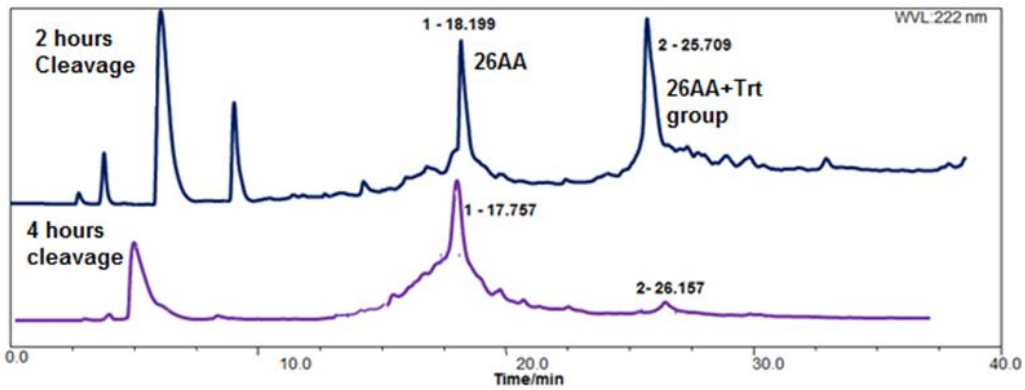
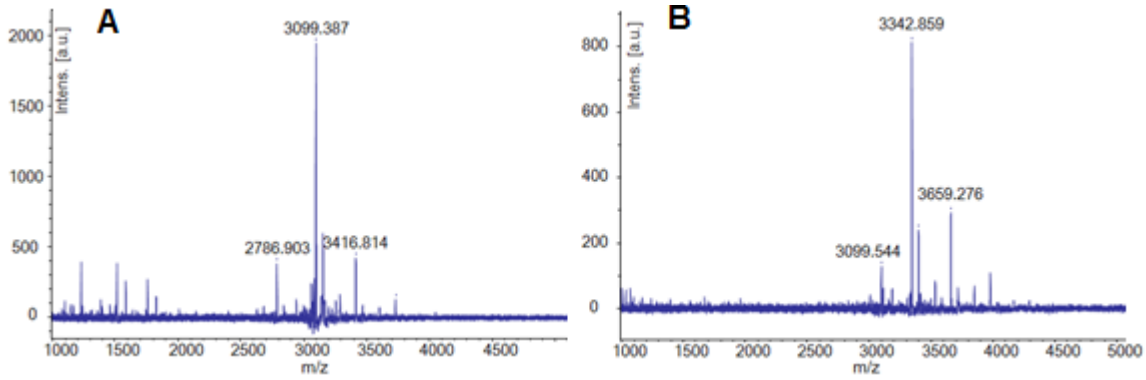


Figure 5 MALDI MS of the two crude major fractions from the analytical C12 HPLC of the 26AA after TFA cleavage. A) the peak at 18th minute, the side chain fully deprotected 26AA; B) the peak at 25.7th minute, 26AA with Cys triphenylmethyl protecting group.



For the purification of the peptide, the C12 column was used instead of the C18 column due to an improved separation of the peptide using the same gradient. Isolated fractions, having been identified by ESI-MS were combined and lyophilised to yield the pure peptide product as a white powder. This was subsequently dissolved in water and analysed using the Ellman's test to determine the concentration of available thiols in solution, which in this case was equivalent to the peptide concentration to calculate the synthesis yield (Table 2).

Table 2 Experimental peptide synthesis yields

	Percentage of Yield
22AA	39%
24AA	41%
25AA	1%
26AA	26%
28AA	26%

22AA and 24AA were obtained in higher yield than the longer 26AA and 28AA peptides. Whereas the low 1 % yield for the 25AA is due to it being an isolated by-product of a faulty synthesis of 26AA, in which the 1st Gly did not couple to the resin.

4.3.3 Characterisation of peptides

Except for the 25AA peptide which was characterised by a very poor quality MALDI MS, and which could not be characterised by ESI-MS due to its low concentration and high salt content, all other peptides were characterised by ESI MS recorded in the positive mode, and studied by C12 analytical HPLC in order to establish purity (Figures 6 and 7).

Figure 6 Analytical C12 HPLC of purified peptides eluted by increasing the gradient from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes followed by a 20 minute wash of 20% acetonitrile (0.05 % TFA). A:22AA, B: 24AA, C: 25AA, D: 26AA, E:28AA. The peak eluting at the second minute in 22AA HPLC is consistent with acetic acid which is used to dissolve the peptide.

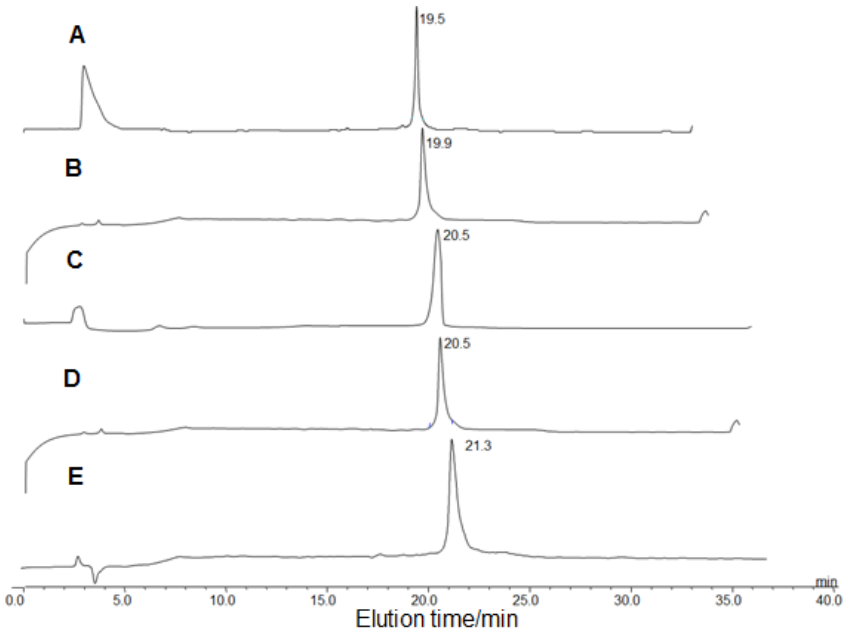
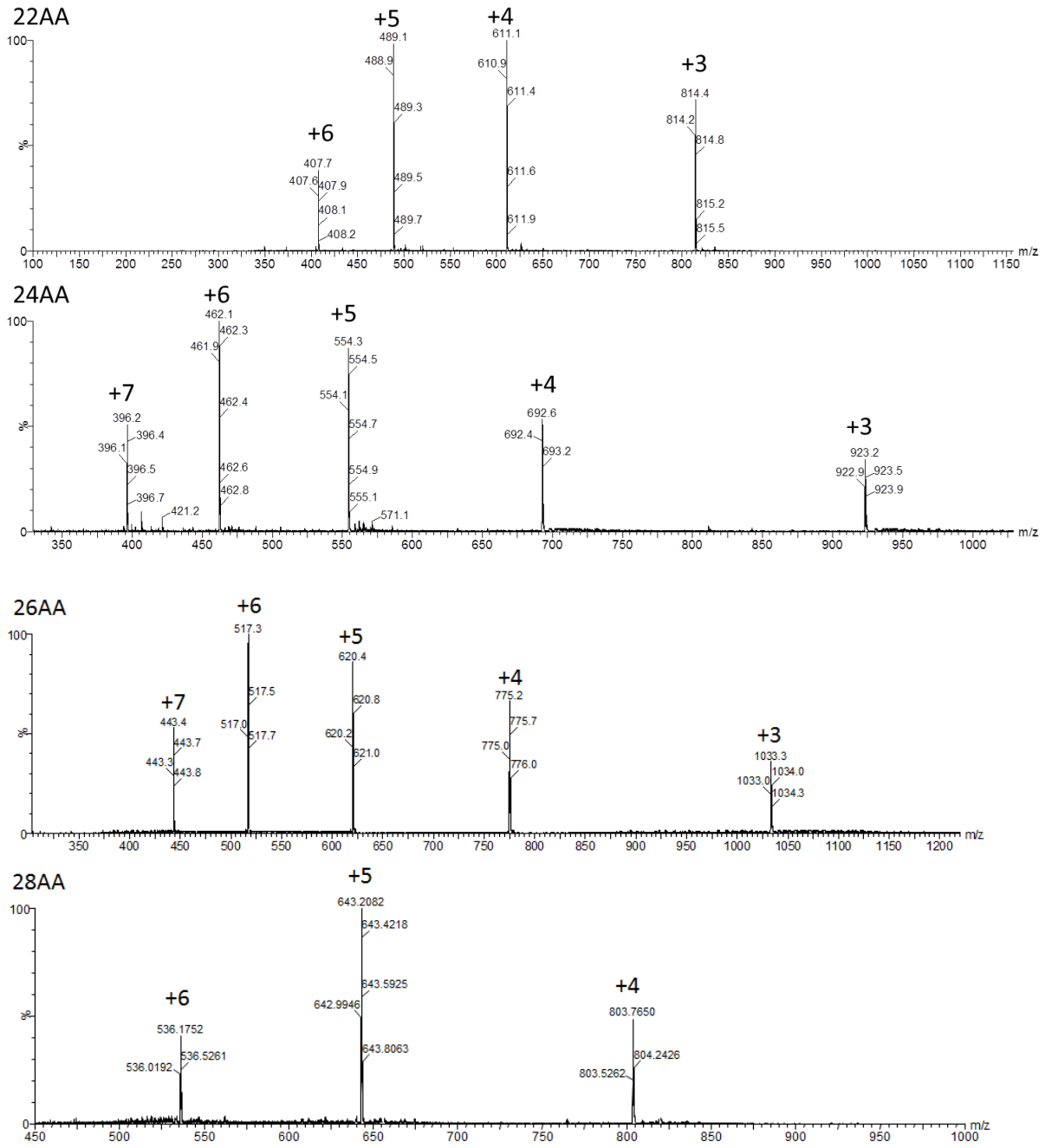


Figure 7 shows the analytical C12 HPLC of the purified peptides by the same gradient. As expected the longer peptides display longer retention times. The 25AA and 26AA peptides share very similar retention times, consistent with their similarity differing only a single Gly located towards C-terminus in the sequence.

Figure 7 ES-MS of 22AA, 24AA, 26AA and 28AA recorded in positive mode.



4.3.4 Disulfide bond formation

Before the successful coupling of designed peptides with 1,1'-bis (hydroxymethylene) ferrocene was developed, the coupling of 28AA was carried out in a glass vial without the protection of a nitrogen atmosphere. Oxidised peptide

dimer formed under these reaction conditions and this was characterised by ES-MS (Figure 8) which depicts the ESI-MS of the fraction collected between 28AA and peptide ferrocene dimer in HPLC. This fraction contains unreacted 28AA peptide monomer, the target dimer and the oxidised dimer. As the oxidised dimer has the molecular weight twice that of 28AA peptide monomer, it can be challenging to assign peaks in the ESI-MS. The oxidised peptide dimer was characterised by the isotopic distributions which confirm the charge state of odd numbers. The addition of dithiothreitol (DTT) reduces the oxidised dimer to peptide monomer and this process was monitored by analytical C12 HPLC (Figure 9). The oxidised peptide dimer peak disappeared 10 minutes after the addition of DTT.

Figure 8 ES-MS of the oxidised peptide dimer recorded 24 hours of the reaction time. +12, +11, +10 and +9 adducts of the peptide dimer were detected; +5 adduct of 28AA was detected; +8, +9 and +11 adducts of oxidised dimer peptide were detected.

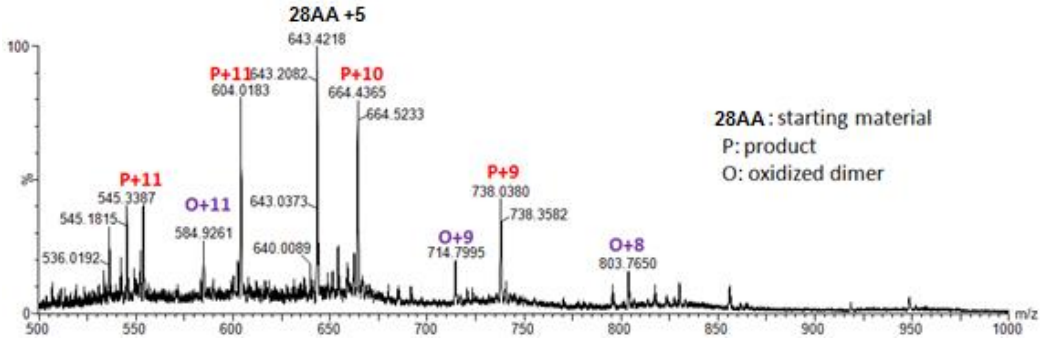
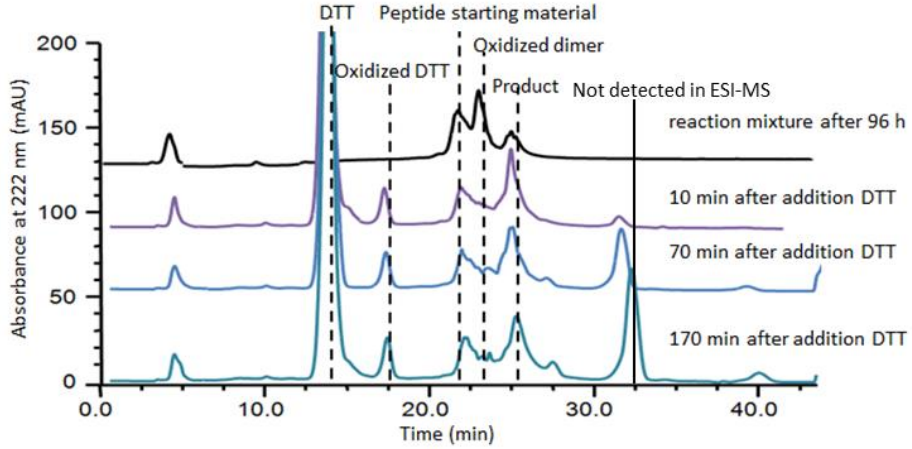


Figure 9 4 analytical C12 HPLC of purified peptides eluted by increasing the gradient from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes followed by a 20 minute wash of 20% acetonitrile (0.05 % TFA). The peak corresponding to oxidised dimer disappeared 10 minutes after the addition of DTT.

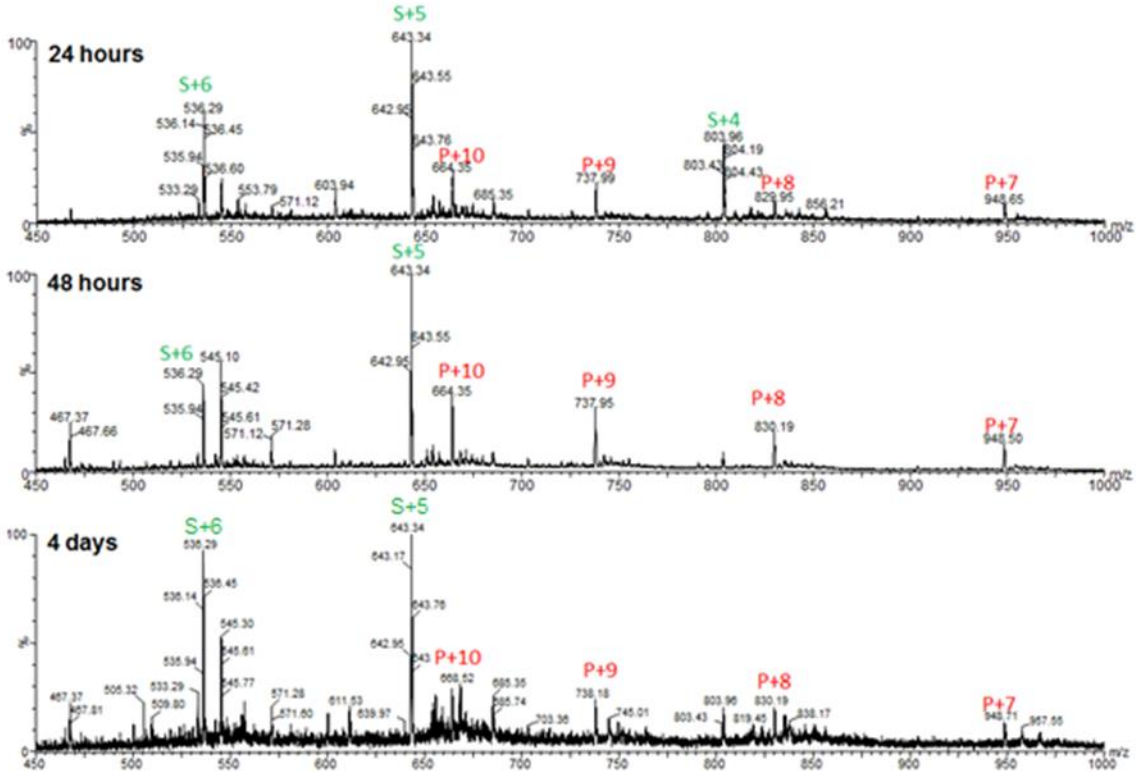


4.3.5 Dimerisation of peptides with 1, 1'-bis (hydroxymethylene) ferrocene

Having successfully synthesised and purified the different peptides based on GCN4, with a single Cys for dimerisation located towards the C-terminus and having studied the dimerisation through Cys with 1,1'-bis(hydroxymethylene) ferrocene (see Chapter 3), we were ready to monitor the coupling reaction between 1,1'-bis(hydroxymethylene) ferrocene and the 28AA peptide.

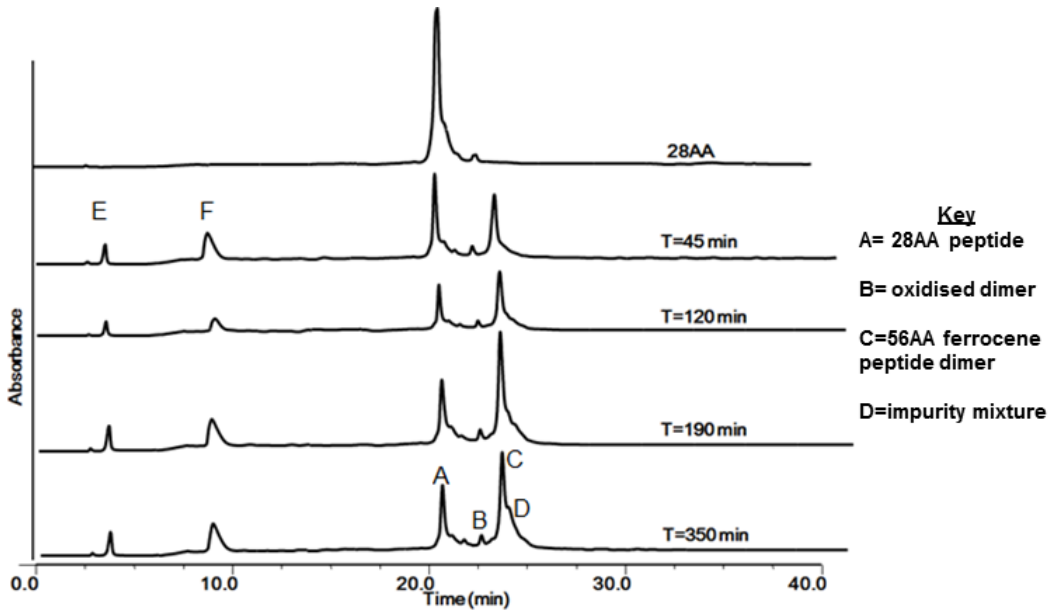
Three equivalents of 28AA were coupled with the ferrocene linker based on the reaction conditions developed in Chapter 3. The progress of the reaction was monitored by both HPLC and ESI-MS. The sample was analysed by ESI-MS after 24 hours, 48 hours and 4 days after the addition of TFA, which is required to catalyse the reaction, see Figure 10.

Figure 10 Positive mode ES MS of 28AA coupling reaction with 1,1'-bis(hydroxymethylene) ferrocene solution monitored after 24 hours, 48 hours and 4 days. +4, +5 and +6 adduct of 28AA were found after 24 hours. S=28 amino acid peptide starting material; P=56 amino acid disubstituted ferrocene peptide dimer.



The ESI-MS of the reaction mixture after 24 hours contains peaks which could be assigned to both the 28AA starting material and the desired product 56AA Fc. After 48 hours both were again detected, though a slightly larger proportion of product. However these two MS spectra can only be interpreted in a qualitative rather than a quantitative fashion. In contrast HPLC can be used to quantitatively monitor progress of the reaction. This required an appropriate HPLC profile to be established that could separate peptide monomer, product dimer and oxidised peptide dimer (an expected side product). The final profile consists of increasing the gradient from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes on a C12 column. After the C12 HPLC gradient had been developed to separate 28AA and 56AA Fc, the reaction between the 28AA peptide monomer and 1, 1'-bis (hydroxymethylene) ferrocene was monitored by C12 analytical HPLC (see Figure 11).

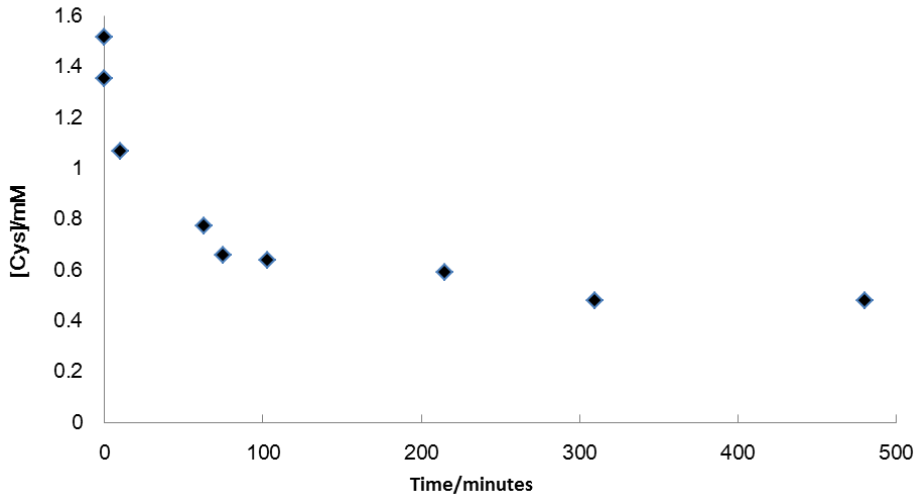
Figure 11 Analytical C12 HPLC monitoring the dimerisation reaction of 28AA with 1,1-bis(hydroxymethylene) ferrocene. Shown are HPLC profiles of a 28AA monomer control and aliquots taken from the reaction at various reaction time. The gradient increases from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes followed by a 20 minute wash of 20% acetonitrile (0.05 % TFA) was used and eluting peaks monitored at 222 nm, No peptide species was detected in the ES-MS of fraction E and F.



In Figure 11, the peak confirmed by ESI-MS assigned to the desired product 28AA Fc dimer steadily increased over 190 minutes in comparison to that assigned to the 28AA peptide monomer. There is no significant change between 190 and 350 minutes reaction time, however the amount of impurity (shoulder after 56AA Fc and labeled D in Figure 9) increased over this time making the dimer more difficult to purify. We were unable to identify the nature of species D by ESI-MS, which contained many m/z peaks including various salt adducts of the product peptide. The small peak between the unreacted 28AA peptide monomer and the desired 56AA Fc product corresponds to the oxidised dimer, confirmed by ESI-MS.

The reaction was also monitored by the Ellman's test which measures the concentration of unreacted Cys thiols from the side chain concentration in the reaction solution (Figure 12).

Figure 12 Concentration of available thiol concentration, equivalent to the concentration of unreacted 28AA peptide, as calculated from the Ellman's test.

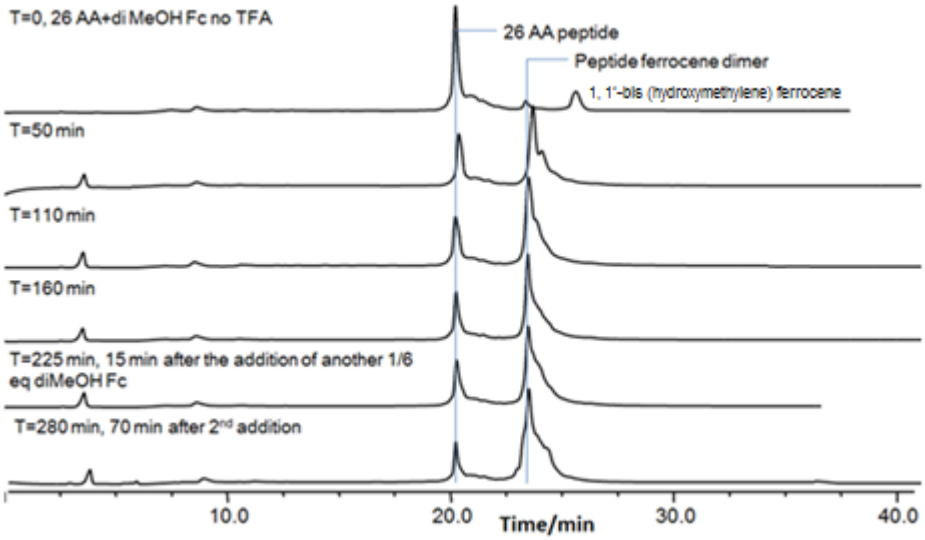


A plot of the thiol concentration as measured using the Ellman's test monitored as a function of reaction time (see Figure 12) shows the concentration of the 28AA starting material decrease over time as the reaction progresses. The two points at 0 min are the concentration of 28AA based on an aliquot from the reaction mixture prior

and post the reaction initiation by addition of TFA. After the first 100 minutes, the concentration of 28AA monomer does not appear to decrease much and thus the reaction has reached equilibrium. The reduction in thiol concentration would be consistent with both product formation (56AA Fc peptide dimer) and the formation of the oxidised peptide dimer as an undesired side product.

The analogous coupling reaction between the 26AA peptide and 1,1'-bis (hydroxymethylene)ferrocene was also monitored by analytical C12 HPLC using the same gradient as developed earlier for the 28AA peptide (Figure 13). The reaction was initially set up with 3 equivalents of the 26AA peptide monomer with respect to ferrocene molecule. However an additional 1/6 equivalent of 1,1'-bis (hydroxymethylene)ferrocene was added into the reaction mixture after 210 minutes of reaction time. A similar change of reaction mixture to that of the 28AA coupling reaction was observed.

Figure 13 Analytical C12 HPLC for monitor the dimerisation reaction of 26AA with 1, 1'-bis (hydroxymethylene) ferrocene. Shown are HPLC profiles of aliquots taken from prior the addition of TFA, the reaction at 50th, 110th, 160th 225th and 280th minute. A gradient increases from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes followed by a 20 minute wash of 20% acetonitrile (0.05 % TFA) was used and eluting peaks monitored at 222 nm, 26AA and product peptide were characterised by ESI-MS. No peptide species was detected in the ES-MS of fraction at 3.5th and 9th minute.



For the coupling of the three remaining peptides 22AA, 24AA and 25AA, three equivalents of peptide monomer was coupled with 1, 1'-bis (hydroxymethylene) ferrocene over 120 minutes, as the reaction mixture appeared to reach equilibrium within 100 minutes and longer reaction times yielded more impurity, resulting in more difficult separation (more peak overlap) of the target peptide dimer products. All the purified ferrocene peptide dimers were characterised by ESI MS and C12 analytical HPLC (Figure 14 and Figure 15).

Figure 14 Analytical C12 HPLC of purified ferrocene peptide dimers eluted by increasing the gradient from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes followed by a 20 minute wash of 20% acetonitrile (0.05 % TFA). A: 44AA, B: 48AA, C: 50AA, D: 52AA, and E: 56AA.

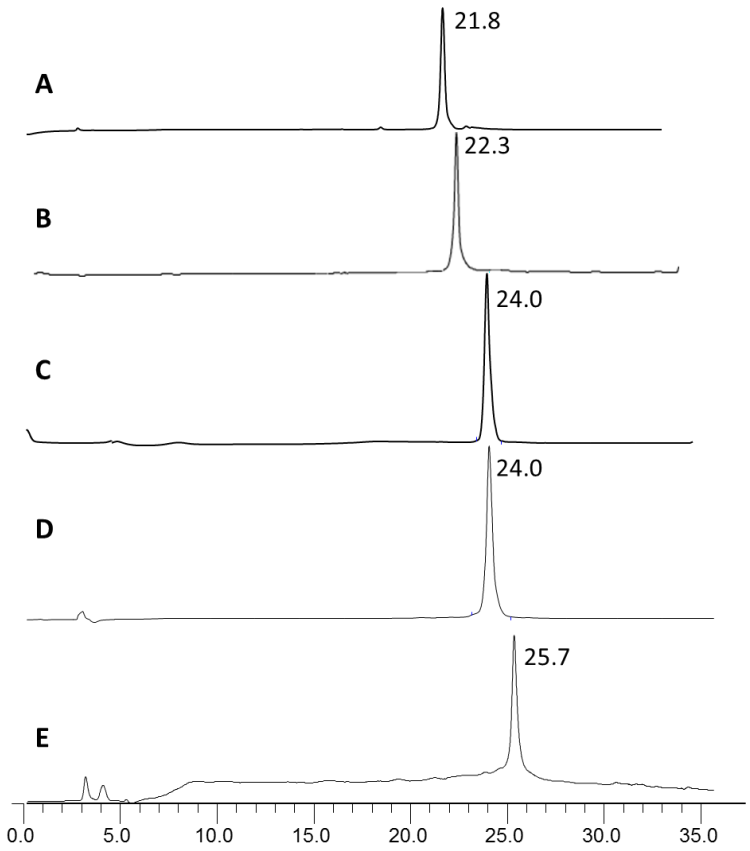
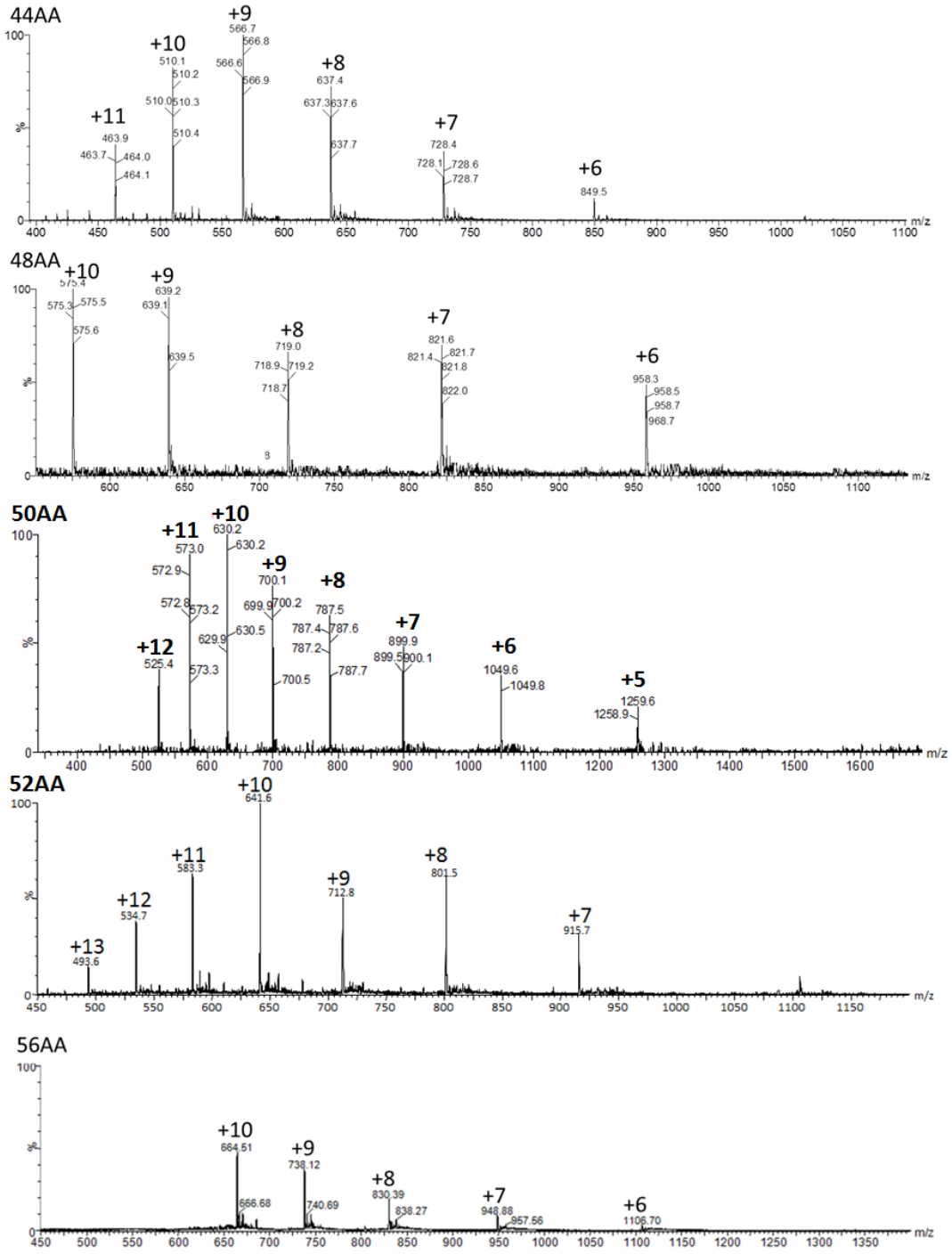


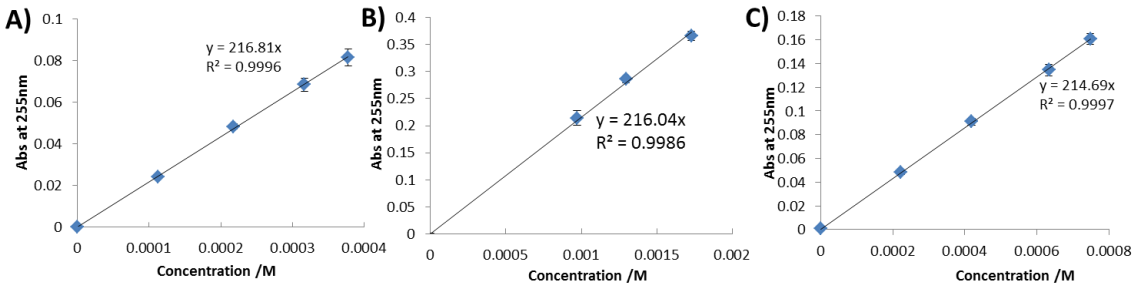
Figure 15 ESI-MS of the 44AA, 48AA, 52AA and 56AA ferrocene peptide dimers recorded in positive mode.



4.3.6 Quantification of ferrocene peptide dimers

The ferrocene peptide dimers were quantified by their UV absorbance at 255 nm and their extinction coefficient was calculated from the sum of the ferrocene extinction coefficient as 255 nm as determined in the previous chapter (Chapter 3) and two equivalents of the peptide extinction coefficient at 255 nm determined here. There is a peak for both 1, 1'-S-ferrocenyldimethyl-L-glutathione and 1, 1'-S-ferrocenyldimethyl-L-cysteine at 255 nm in their UV spectra. These two small ferrocene peptide molecules share the same extinction coefficient at 255 nm. The average value of 10800 L mol^{-1} cm^{-1} was used as the extinction coefficient of our ferrocene linker domain. This peak was also found in all the UV-visible spectra recorded of all the ferrocene peptide dimers recorded under similar conditions. In order to calculate the extinction coefficient of the peptide, the UV visible spectra of the 24AA, 26AA and 28AA peptides were recorded at various concentrations. These experiments were repeated in triplicate in order to validate the data (Figure 16).

Figure 16 Extinction coefficient calculations of A) 24AA, B) 26AA and C) 28AA peptide monomers at 255 nm. The average absorbances at 255 nm are plot as a function of peptide concentrations. The error bars shown corespond to the standard deviation of the data recorded in triplicate.



Extinction coefficient determination for peptide contribution was performed in triplicate. The average absorbance at 255 nm was plot as a function of peptide

concentration. Standard deviation was used as the error bar. These three peptides share similar extinction coefficients at 255 nm, and so the average value of 216 $Lmol^{-1}cm^{-1}$ was used for subsequent calculation for ferrocene peptide dimers, in which an extinction coefficient of 11232 $Lmol^{-1}cm^{-1}$.

4.3.7 Circular dichroism (CD) of peptide monomers and ferrocene peptide dimers with CRE and non-specific DNA

The GCN4 basic domain is unstructured in the absence of DNA containing a target site, and this therefore provides us with a method by which to test the peptide dimers' binding to different DNA sequences by monitoring their folding.[127] In order to compare the structures of the peptide monomer and ferrocene peptide dimers in the absence and presence of CRE and non-specific DNA, circular dichroism (CD) was carried out on solutions containing either 10 μM peptide monomer or 5 μM peptide dimer in the absence and presence of 5 μM DNA.

As described, we designed peptide sequences with different hinge domains and subsequently different distances between the DNA binding basic domain and the dimerisation unit (ferrocene). This will allow us to study the spatial relationship between the two domains and the influence this will have on their DNA binding affinity, see Table 3. In the shorter sequences, the 22AA and 24AA peptides only have one or three residues, respectively, in the hinge domain, which should allow us to evaluate what is the shortest hinge required for DNA binding. There is one Gly located at the C-terminus of 26AA which is not present in 25AA. This should allow us to evaluate if sequences extending beyond the C-terminus play an important role. Both 26AA and 28AA sequences contain the whole hinge, differing only in two Gly

residues as a spacer between the peptide basic domain and the dimerisation unit. These two sequences were designed to study if a flexible spacer between the peptide and the ferrocene dimerisation unit, affects their DNA binding affinity.

Table 3 Peptide sequences used in these studies

Peptide	N-terminus	Basic domain	Hinge	C-terminus
22AA	AL	KRARNTEAARRSRARK	L	GCG
24AA	AL	KRARNTEAARRSRARK	LQR	GCG
25AA	AL	KRARNTEAARRSRARK	LQRMKQ	C
26AA	AL	KRARNTEAARRSRARK	LQRMKQ	CG
28AA	AL	KRARNTEAARRSRARK	LQRMKQ	GGCG

4.3.7.A CD of peptide monomers with CRE and non-specific DNA

All five peptide monomers were studied by CD to establish folding. Solutions of 10 μM peptide monomer in the presence and absence of 5 μM duplex DNA, and in the presence of 100 mM NaCl and 10 mM phosphate buffer pH 7, chosen to mimic physiologically relevant conditions, were studied by CD (see Figure 17).

A very noisy signal was obtained for the CD of 22AA peptide using 10 scans and so the scan number was increased to 16. This gave a reasonably smooth CD spectrum. The CD spectra of the peptide monomers could be consistent with a 3_{10} helix, and thus they are either in random coil or 3_{10} helical in the absence of target DNA,[128-130] but form α-helixes in the presence of CRE-DNA. The residual molar ellipticity, $[\theta]_{RME}$, of the peptides was calculated from the CD, and the residual molar ellipticity at 222 nm was used to calculate the percentage of α-helical folding, which was found to be 16% for 22AA peptide in the absence of DNA, see Table 4.[131,132] Addition of DNA containing either the CRE target site or a NS strand, showed no significant change in the folding.

Figure 17 CD spectra of 10 μM peptide monomers in the absence and presence of 5μM CRE and NS DNA. Independent experiments were performed in triplicate. CD of A), B) and C) 22AA; D), E), and F) 24AA, G), H) and I) 25AA, J) K) and L) 26AA, M), N) and O) 28AA in 100 mM phosphate pH 7 and 10 mM NaCl. A) is the CD of 22AA by 10 scans, and it is not used for the evaluation of data. NS DNA sequence: 5'-TGGAGTATGCGTCGATTCGT-3'

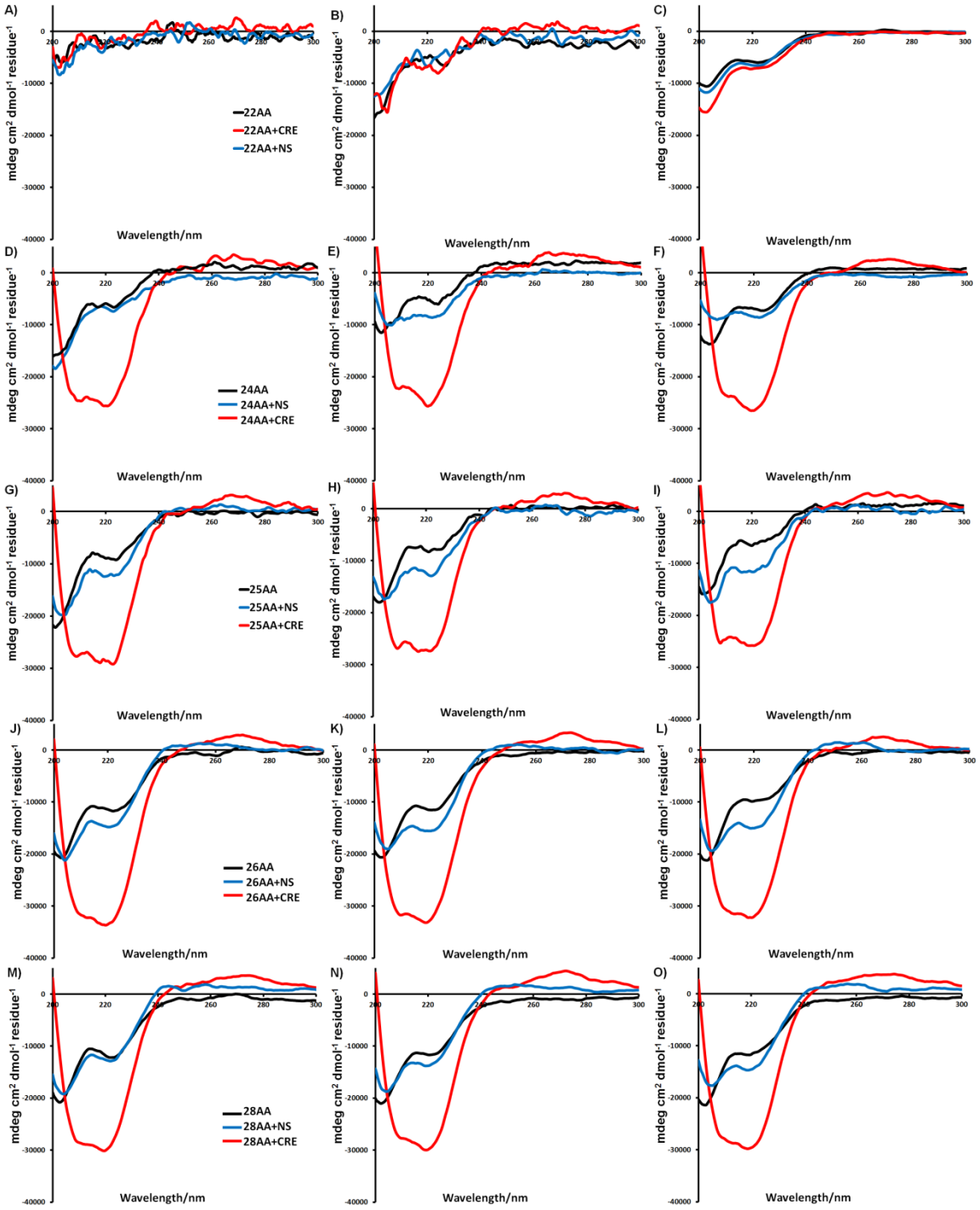


Table 4 The average percentage of α-helical content of peptide monomers in the absence and presence of CRE and non-specific (NS) DNA, based on $[\theta]_{222}$ from CD. The CD of 22AA was only repeated in twice, and thus no standard deviation has been calculated.

peptide monomer	Alone	with CRE DNA	with NS DNA
22AA	16	20	17
24AA	19±3	70±2	21±3
25AA	22±4	78±3	34±5
26AA	30±3	90±2	42±2
28AA	33±2	82±4	38±2

A trend can be observed between increasing the peptide length and greater folding, see Table 4. All peptide monomers, except the shortest 22AA peptide, display intense negative minima at 208 and 222 nm in the CD spectra recorded in the presence of DNA containing the CRE target site. However analogous experiments recorded in the presence of non-specific (NS) DNA resulted in a significantly smaller change to the CD spectra, see Figure 17. The minima at 208 and 222 nm are consistent with formation of an α-helix, and therefore sequence-specific DNA binding. Notably the 26AA peptide is the most folded in the presence of DNA, suggesting there is no advantage in the two additional Gly residues in the 28AA peptide.

Following these experiments in the presence of CRE DNA, the solutions were investigated by ESI-MS. All peptides were detected as monomers, confirming that they did not dimerise through formation of a disulfide bond through the Cys located towards their C-terminus. Therefore changes in the CD are truly due to binding of peptide monomers to the DNA target site. The binding of GCN4 peptide monomers with DNA, has previously been reported in the literature.[120-122] Peptide monomers based on GCN4, which lack a dimerisation unit, have also been reported to recognise the DNA CRE target site. Lai and co-workers designed a peptide based on GCN4, and this sequence contains the whole hinge domain and the basic region.[122]

Lai's peptide monomer is 10 % and 70 % folded respectively in the absence and presence of CRE DNA, which is similar to that reported for peptide dimers in which the GCN4 basic domains are dimerised by a disulfide bond.[133] Lai's peptide sequence is similar to our 25AA and 26AA sequence except for the Cys and Cys-Gly cap located at our C-terminus. However, all our peptide monomers are more folded both in the absence and presence of CRE DNA. This suggests that the Cys or the Cys-Gly cap at the C-terminus stabilises the formation of the α-helix in the presence and absence of CRE DNA. The stabilisation of peptide α-helixes by a terminal Cys has previously been reported in the literature.[134,135] In the presence and absence of DNA, the 26AA monomer is more helical than the analogous 25AA and has the highest helicity of all our peptides. The 26AA peptide has a C-terminal Cys-Gly while the 25AA peptide has only a single Cys at the C-terminus. It has been reported that the preference for N-terminal Gly in the α-helix, while C-terminal Gly cannot stabilise the α-helix, but does frequently occur in α-helical motifs.[133,136] This has not previously been investigated.[136] The 28AA peptide has a lower helicity than the 26AA peptide which suggests that the two Gly spacers between the hinge and the C-terminal Cys-Gly may interfere with any Cys stabilisation.

4.3.7.B CD of ferrocene peptide dimers with CRE and non-specific DNA

As we now have evidence that these peptide monomers are able to display sequence-specific DNA binding, it remained for us to evaluate the ferrocene peptide dimers, in order to establish if the ferrocene dimerisation domain would be tolerated and if sequence specific DNA binding could be retained.

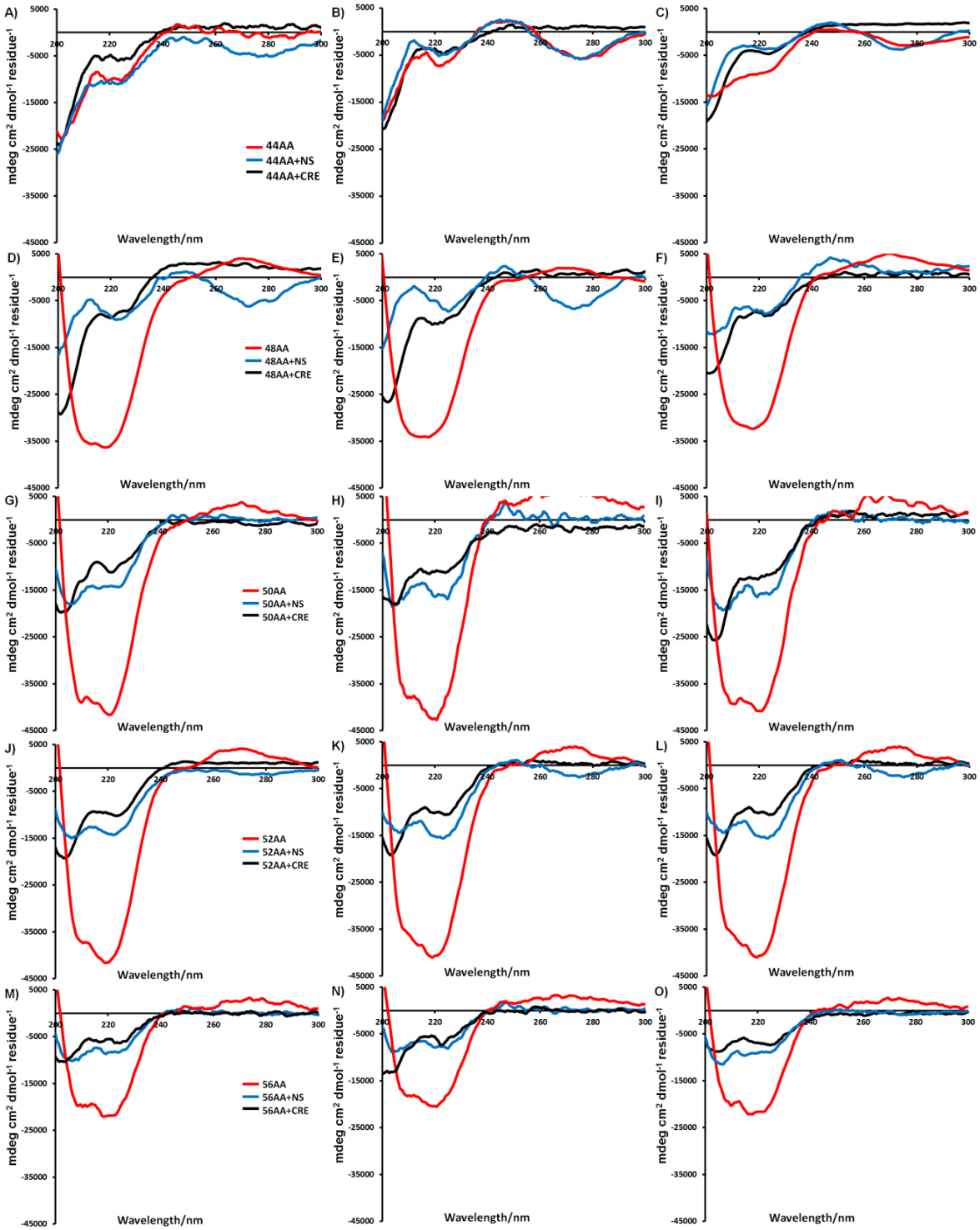
All CD experiments with ferrocene peptide dimers were performed at 5 μM, in the

absence and presence of 5 μM DNA by recording and averaging 16 CD scans. Experiments were repeated in triplicate in order to validate the data, see Figure 18. The molar ellipticity and the percentage folding were calculated by the method mentioned in the previous section, see Table 5.[131] Again the percentage folding trends to increase as the number of residues increases, except for the 56AA dimer, which suggests that two Gly spacers introduced between the hinge and the Cys may interfere with any peptide stabilisation by Cys, similar to that observed for the peptide monomers.

Table 5 The average folding percentage of ferrocene peptide dimers in the presence and absence of DNAs, based on [θ]222 from CD.

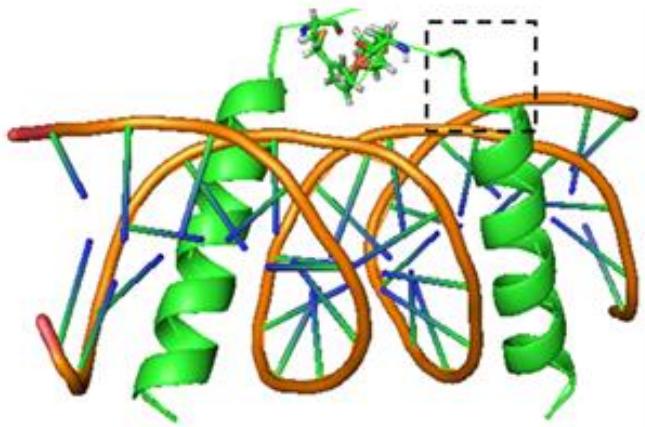
Peptide dimers	alone/%	Peptide with NS DNA/%	Peptide with CRE DNA/%
44AA	14±2	19±2	24±2
48AA	24±1	22±2	92±3
50AA	32±1	42±2	115±2
52AA	29±2	38±1	113±3
56AA	18±1	22±1	55±2

Figure 18 CD spectra of peptide dimers in the absence and presence of CRE and NS DNA. Experiments were repeated in triplicate. CD of A), B) and C) 44AA; D), E), and F) 48AA, G), H) and I) 50AA, J) K) and L) 52AA, M), N) and O) of 56AA in 100 mM phosphate pH=7 and 10 mM NaCl.



There is only a small increase in the helicity for the 44AA ferrocene peptide dimer in the presence of DNA, which indicates that it cannot bind to DNA efficiently. Presumably this is because the single residue hinge domain is unable to provide the minimum length required between the basic domain and the ferrocene dimerisation domain, in order to bind DNA as a helical dimer. These findings are supported by our molecular dynamics (MD) simulations up to 10 ns (see Figure 19). The MD model of 44AA-CRE complex clearly shows that one of the peptide strands is unwinding. The 44AA ferrocene peptide dimer is therefore unable to bind to the CRE DNA target sites with the basic domains folded into α-helixes, which is important for strong and selective DNA binding.

Figure 19 The MD simulation model of the complex between the 44AA and CRE DNA at 10 ns. The box in the figure shows the unwinding of one of the peptide strands. The simulation starting model was modified from the 2DGC PDB file a GCN4bZIP-CRE DNA complex.[29]

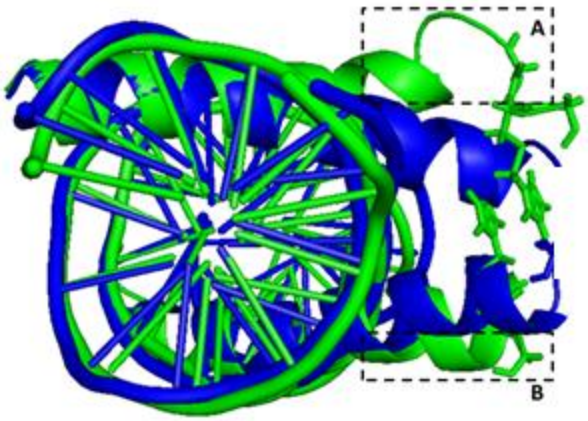


The 48AA, 50AA, 52AA and 56AA ferrocene peptide dimers display higher helicity in the presence of CRE DNA than their monomer analogues. All these dimers bind to DNA sequence selectively. This is due to dimerisation by the ferrocene unit which provides the necessary “free” rotation of the two peptide strands required in order to correctly orientate the basic region and therefore stabilise the resulting ferrocene peptide dimer-CRE DNA complex. It is in good agreement with the reported

literature, in which the dimerised peptide based on GCN4 can bind to the DNA target site while the isolated peptide monomer fails to do so with a high affinity.[35]

The 48AA ferrocene peptide dimer is less folded and appears to have less binding affinity than either 50AA or 52AA with CRE DNA. This is supported by the MD simulations in which one strand of 48AA is beginning to unwind and both strands are pushed away to maintain the 48AA ferrocene peptide dimer-CRE DNA complex model. Thus the short hinge of the 48AA peptide sequence is not long enough for the optimal spatial relationship between basic domain and ferrocene dimerisation unit for CRE DNA binding (see Figure 20).

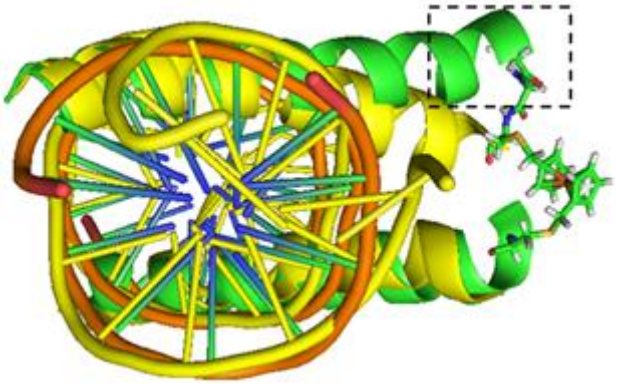
Figure 20 The overlay of the MD simulation models of the ferrocene peptide dimer-DNA complex after 10 ns. 52AA-DNA complex in blue (the ferrocene unit is not shown); 48AA-DNA complex in green. One 48AA peptide strand unwinds in box A while both strands are pushed away for the CRE DNA (in box A and B). The simulation starting model was modified from the 2DGC PDB file of a GCN4-bZip CRE DNA complex.[29]



CD of the 56AA ferrocene peptide dimer is consistent with selective DNA binding, however it is less helical than the 28AA peptide monomer, with apparently reduced binding affinity to CRE DNA than 28AA. In Figure 21, the simulation model of the 56AA ferrocene dimer-CRE DNA complex is overlaid with that for the 52AA ferrocene

dimer-CRE DNA complex which has the highest DNA binding affinity as determined by CD in the four simulated ferrocene peptide dimers. One of the peptide strands of the 56AA ferrocene peptide dimer is pushed away from the CRE DNA compared to the model with the shorter 52AA ferrocene peptide dimer, and thus the two C-terminal Gly spacer appears to hinder the formation of optimal conformation for correct alignment of the two α-helices of the 56AA ferrocene peptide dimer, for effective binding to the CRE DNA target site.

Figure 21 The overlay of the MD simulation models of the ferrocene peptide dimer-DNA complex after 10 ns. 52AA ferrocene dimer-CRE DNA complex in yellow (the ferrocene unit is not shown); 56AA ferrocene dimer-CRE DNA complex in green. One 56AA peptide strand is pushed away for the CRE DNA in box. The simulation starting model was modified from the 2DGC PDB file of a GCN4-bZip CRE DNA complex.[29]



The only difference between the peptides sequences in 50AA ferrocene peptide dimer and 52AA ferrocene peptide dimer is the single Gly located beyond the Cys for dimerisation towards the C-terminus. They behave similarly in the CD experiments, which indicates that they bind to DNA selectively and have the highest DNA binding affinity of the various dimers investigated. In the MD simulation of 52AA-CRE DNA complex, the two peptide strands are not unwound or pushed away from the CRE DNA binding site and are more folded than for the 44AA, 48AA and 56AA ferrocene peptide dimers, supporting its higher DNA binding affinity, see Figures 19-21.

However, the 50AA and 52AA ferrocene peptide dimers are both calculated as being more than 100 % folded in the presence of CRE DNA based on CD ellipticity at 222 nm. This change must be due to the change of peptide helicity, as well as another conformational change on formation of the peptide-CRE DNA complex. In the CD spectra of the 48AA, 50AA, 52AA and 56AA ferrocene peptide dimers with CRE DNA, a positive induced CD signal appears at 275 nm, which can be assigned to a protein-induced DNA bending of B-type DNA.[136-138] For wild type GCN4, it must bend and distort the CRE site to maintain the important basic region-phosphate interactions, and the critical Thr236-phosphate backbone contact.[29,139] Mascarenas' azobenzene linked peptide dimer based on GCN4 also bends the DNA when it binds to CRE DNA in the cis conformation.[36] We compared the CRE model from GCN4 bZip-CRE DNA complex crystal structure in the literature with our simulation model of the 52AA ferrocene dimer-CRE DNA complex, and can observe a bending towards the peptide dimerisation unit, see Figure 22.

Figure 22 The overlay of CRE DNA in the crystal structure of the complex formed with native GCN4 (green) and the CRE from the MD simulation model of the 52AA ferrocene dimer-CRE DNA complex at 10ns (blue). Bending of the latter towards the ferrocene dimerisation domain can be observed. The simulation starting model was modified from the 2DGC PDB file of a GCN4-bZip CRE DNA complex.[29]

In our peptide design for the 50AA and 52AA ferrocene peptide dimers, we have included both the basic and hinge domains of native GCN4, and thus the stronger 50AA and 52AA ferrocene dimer DNA bending than native GCN4, must be due to the

ferrocene dimerisation unit and the short Gly-Cys located at the C-terminus. We measured the distance between the two sulphur atoms of Cys side chains in the 52AA ferrocene dimer-CRE DNA MD model to be 4.5 Å (see Figure 23), which is much shorter than the 8 Å distance between the side chains on the two Cys residues, which replace the first residue in the leucine zipper domain in the crystal structure of the GCN4-CRE DNA complex (see Figure 2). Therefore, the ferrocene dimerisation domain holds the two peptide strands closer together than in the native GCN4 bZip complex, subsequently bending the CRE DNA more.

Figure 23 Distance measured between the sulphur atoms of the two Cys residues in the 52AA ferrocene dimer-GCN4 DNA MD model. The simulation starting model was modified from the 2DGC PDB file of a GCN4-bZip CRE DNA complex.[29]

4.3.8 Electrochemical characterisation of ferrocene peptide dimers

In the previous chapter, we reported the electrochemistry of our precursor linker molecule, 1,1'-bis (hydroxymethylene)ferrocene and the product of two short peptides (GSH) dimerised through their Cys side chain, 1,1'-s-ferrocenyldimethyl-L-glutathione (diGSFc). 1,1'-Bis (hydroxymethylene)ferrocene displayed fully reversible electrochemistry consistent with a single electron transfer process by the

experimental set up with all three electrodes emerged in the sample solution, while diGSFc did not display completely reversible electrochemistry by experimental set up with working electrode in a sample holder.

We therefore aimed to evaluate if the attractive electrochemical properties of ferrocene were retained on linking to larger peptide units based on GCN4. The electrochemical characterisation was carried out in two experimental kits. In kit A, the ferrocene peptide dimer was in a narrow-bottom conical glass vial with a stirrer bar to raise the liquid level for adequate immersion of electrodes, and all the electrodes are emerged in the sample solution. In kit B, the ferrocene peptide dimer was in a 200 µL sample holder with working electrode, the sample holder was submerged in a 20 mL sample vial with reference and counter electrodes held in a 0.1 mM phosphate buffer pH=7 and 0.1 M NaCl solution, see Chapter 2 Figure 6.

4.3.8.1 Differential Pulse Voltammetry titration of 52AA Fc by Set A

Differential pulse voltammetry (DPV) is more sensitive than cyclic voltammetry, so it can detect ferrocene peptide dimers at lower and more biological relevant concentrations. 52AA was first characterised by DPV to test the experimental kit as it required significantly less sample.

Table 6 DPV titration of 52AA Fc

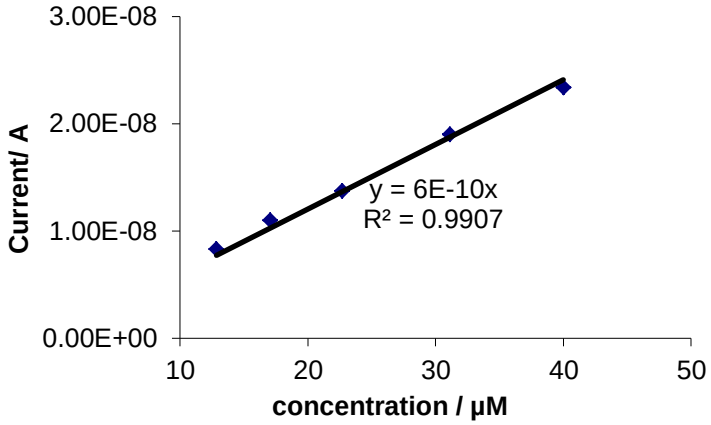
Concentration (µM)	Potential/ V	Current/ A
40	0.314	2.34×10^{-8}
31.1	0.313	1.90×10^{-8}
22.7	0.313	1.37×10^{-8}
17.0	0.313	1.10×10^{-8}
12.9	0.313	8.31×10^{-9}

The titration of increasing concentration of the 52AA ferrocene dimer, the peak

potential remains the same (table 6) and the peak current is proportional to the concentration of peptide dimers (Figure 24), and thus this experimental set up gave valid data for DPV.

The ferrocene peptide dimer can stick to the gold electrode surface. The DPV of used electrode after 52AA DPV in blank buffer solution still display a peak at 0.313 V, and this peak disappeared after the electrode was polished and cleaned.

Figure 24 DPV titration of 52AA peptide dimer, the peak currents are plotted against the concentration of ferrocene peptide dimer in 0.1 M NaCl and 0.01 M phosphate buffer pH=7.



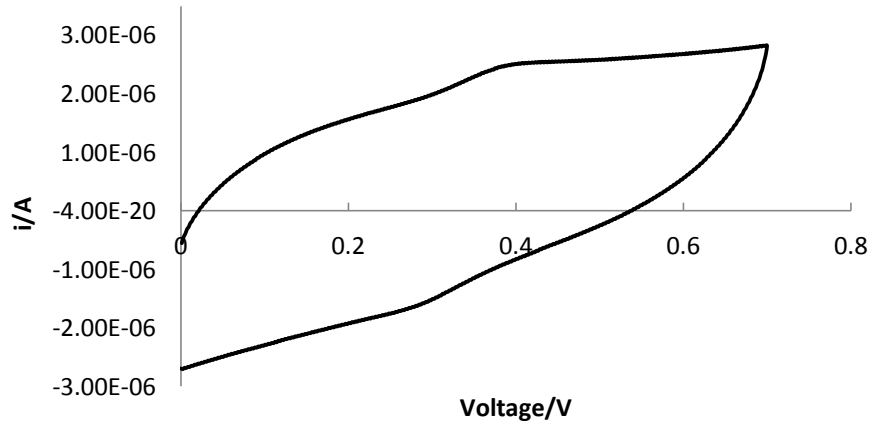
4.3.8.2 Cyclic voltammetry of 52AA Fc by Set A

The previous section demonstrated a valid DPV of 52AA ferrocene peptide dimer. We now moved on to the cyclic voltammetry (CV) study of the ferrocene peptide dimer. CV can provide us more information about the electrochemical properties of the ferrocene peptide dimers as CV includes a reversed scan in which both the oxidation and reduction of the ferrocene peptide can be studied while only an oxidation process was recorded in DPV.

The cyclic voltammetry (CV) experiment of 52AA Fc by set A was conducted using both a carbon and a gold working electrode. 0.1 mM 52AA Fc was scanned by CV at

various different scan rates to check the reliability of the experimental data obtained using set A. These experiments were performed in triplicate. In the CV recorded using a carbon working electrode, the reduction peak is too weak to be processed out (Figure 25).

Figure 25 CV of 0.1 mM 52AA Fc by carbon eletrode at 200 mV/s using carbon working electrode.

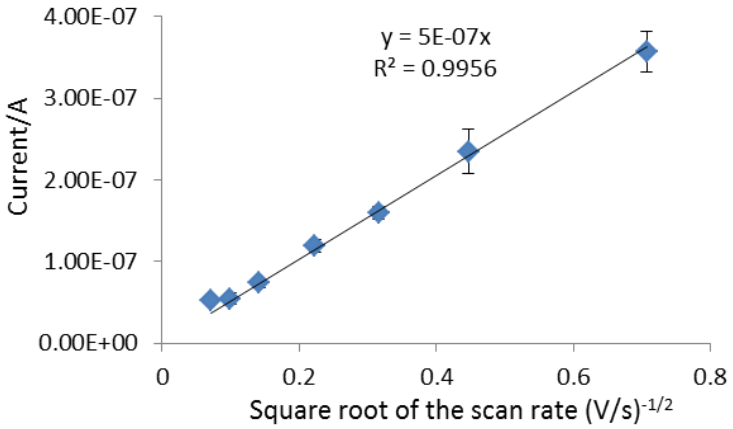


Under these conditions, the ΔE_p values are all much larger than 59 mV, which is not consistent with a one electron redox process. This could due to the resistance of the sample holder frit where the electron transfer occurs between the sample solution in the holder and the buffer solution in the vial. In the CV of diGSFc by the same experimental kit, they all displayed a ΔEp around than 100 mV which cannot indicate the one electron process in CV, so it may be due to the frit of sample holder.

In Cottrell equation, i_{pa} and i_{pc} are proportional to the square root of the scan rate ($t^{-1/2}$). The average i_{pa} was plot as a function of the square root of the scan rate. (Figure 26) The error bar shown correspond to the standard deviation of the data recorded in triplicate. Except for the i_{pa} at 5 mV/s, the trend line fully within the error bar of all average ipa, therefore 5 mV/s is too slow to provide validate CV data for the 52AA

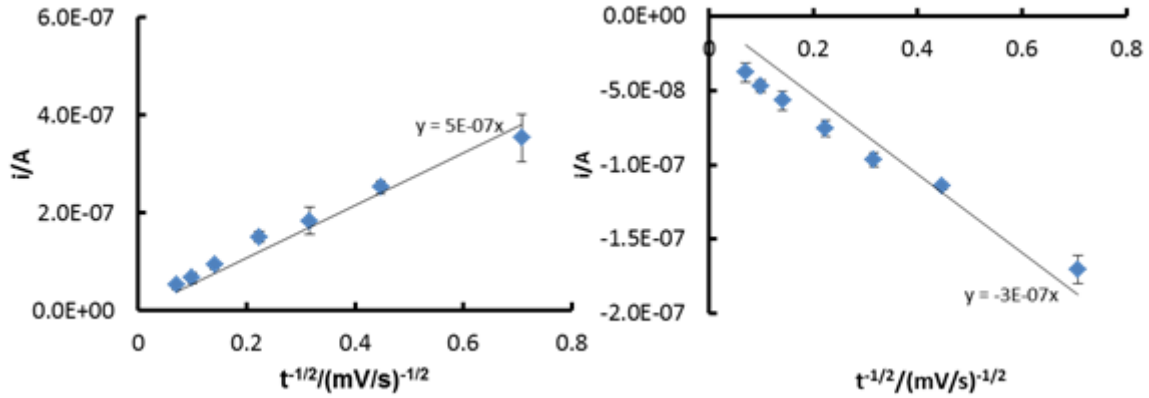
peptide.

Figure 26 CV of 52AA Fc by carbon electrode, the average ipa was plotted against the square root of the scan rate. The error bar shown corespond to the standard deviation of the data recorded in triplicate.



0.1 mM 52AA was scanned by CV using gold working electrode at variant scan rates to check its reliability by plotting i_{pa} and i_{pc} against the square root of scan rate. This experiment was repeated in triplicate to validate the data.

Figure 27 CV of 0.1 mM 52AA by gold electrode, average ipa and ipc was plotted as a function of the square root of scan rate. The error bar shown corespond to the standard deviation of the data recorded in triplicate.



In Figure 27, the trend lines of i_{pa} and i_{pc} are outside most of the error bar of average value, and thus the gold electrode CV of 52AA by kit A is not reliable.

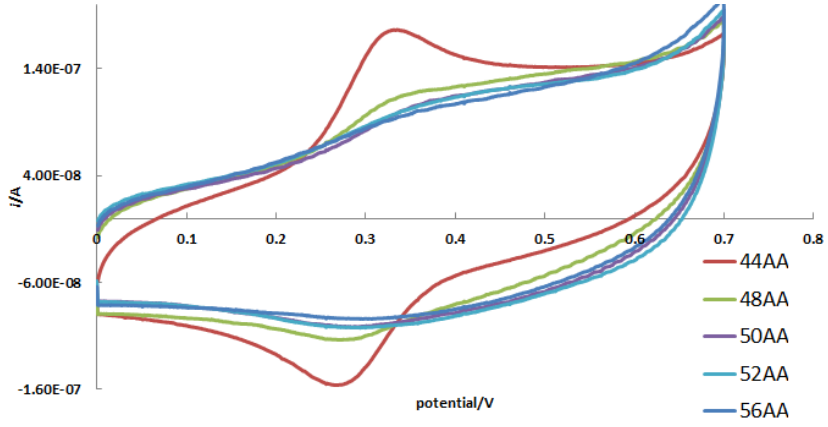
4.3.8.3 Cyclic voltammetry of ferrocene peptide dimers by Set B

As the CV of 52AA peptide by kit A cannot yield reliable data, we continued the CV experiments using kit B in which all the electrodes are emerged in the sample solution, so the sample holder frit cannot interfere the electron transfer process.

We recorded the CV of all the ferrocene peptide dimers at 100 mM concentration and with a scan rate 20 mV/s. Each CV experiment was performed in triplicate using fresh sample stock solutions.

The 44AA peptide dimer has the shortest peptide strand and larger i_{pa} and i_{pc} than other dimers, see Figure 28. Its ferrocene redox center is more exposed to the solution for electron transfer. This implies that electrochemical activity of the redox center can be affected by the steric hindrance from the presence of molecule linked to the redox center by preventing charge transfer to and from the electrode. Except for 48AA, the ferrocene peptide dimers with longer peptide sequences have lower i_{pa} and i_{pc} values due to the steric hindrance of the peptide's random conformation. Similar observation have previously been reported by Jenkins and co-workers.[83] In their study, the presence of a DNA strand attached to a ferrocene shields the ferrocene resulting in a less intense current in the DPV.[83]

Figure 28 CV of the ferrocene peptide dimers (100mM) recorded at 20 mV/s using gold working electrode in experimental kit B.



In the CV of the peptide dimers, their ΔE_p are all around 59 mV which indicate a one electron redox process under these conditions; except for 56AA, their $|i_{pa}/i_{pc}|$ are all 1.01 or 1.02 which indicates that 44AA, 48AA, 50AA and 52AA peptides all display fully reversible for an oxidation of the ferrocene group to the ferrocenium cation. 56AA has the lowest i_{pa}, i_{pc} and a $|i_{pa}/i_{pc}|$ of 1.08 and therefore gives the least electrochemistry. This could be because of its longer peptide in random conformation without DNA which can shield the ferrocene redox center from the access of the counter ions. 50AA, 52AA and 56AA share similar E_{pa} and E_{pc}, because they have similar sequences except for the C-terminal Gly spacer. 44AA and 48AA are more cathodic and easier to be oxidised than the longer ferrocene peptides dimers. A possible explain is that they do not have the entire hinge and both have one less Lys residue which is positively charged (see Table 7).

Table 7 The average analytical data of the ferrocene peptide dimers. $\Delta E_p = E_{pa} - E_{pc}$; $E_{1/2} = (E_{pa} + E_{pc})/2$. All the peptide dimers' $|i_{pa}/i_{pc}|$ are around 1 and all the peptide dimers' ΔE_p are around 59 mV

	i_{pa}/A	i_{pc}/A	E_{pa}/mV	E_{pc}/mV	$\lvert i_{pa}/i_{pc} \rvert$	ΔE_p/mV	$E_{1/2}$/mV
44AA	8.70E-08±3.48E-09	-8.59E-08±-4.30E-09	329±8	271±7	1.01	58	300
48AA	1.74E-08±7.48E-09	-1.73E-08±-6.40E-09	320±5	262±6	1.01	58	291
50AA	1.97E-08±8.47E-09	-1.95E-08±-6.65E-10	358±6	301±4	1.01	57	330
52AA	1.90E-08±9.69E-10	-1.87E-08±-3.93E-10	357±8	297±6	1.02	60	327
56AA	1.29E-08±3.74E-10	-1.20E-08±-1.74E-10	357±9	300±4	1.08	57	329

4.4 Conclusions

Five different peptide sequences based on GCN4 were designed, synthesised, characterised and purified. The longer peptide displayed a longer retention time in HPLC. These peptides were coupled to 1,1'-bis (hydroxymethylene)ferrocene to yield the desired ferrocene peptide dimers. The coupling reaction was monitored by the

Ellman's test and analytical HPLC, and required much less time than the coupling of the short peptide glutathione with 1, 1'-bis (hydroxymethylene) ferrocene. The lack of a nitrogen atmosphere for the coupling reaction resulted in the formation of oxidised peptide dimer, which could be reversibly reduced by the addition of excess DTT. The ferrocene peptide dimers were quantified through their absorbance at 255 nm in their UV visible spectra, and the extinction coefficient was calculated by the sum of the small ferrocene molecules' extinction coefficient and two equivalents of the extinction coefficient of the peptide all at 255 nm.

The peptide monomers and ferrocene peptide dimers have been studied by CD. All the peptide monomers display sequence specific binding to DNA based on CD spectroscopy, except for the 22AA peptide monomer which is too short for the recognition of target DNA, presumably as it does not retain sufficient conserved residues from native GCN4.

The ferrocene peptide dimers bind to DNA, based on CD studies, except for the short 44AA ferrocene peptide dimer, the sequence of which is again too short for recognition of target DNA while its hinge domain sequence is too short to maintain the spatial relationship between the DNA binding basic domain and the ferrocene dimerisation unit. These findings are supported by MD simulations. The peptide monomers are less helical in the presence of CRE DNA, consistent with reports that their binding is weaker. It is because of dimerisation, achieved by a leucine zipper in native GCN4 and by the ferrocene unit described here. The ferrocene unit provides the necessary flexible "free" rotation of the two peptide strands required in order to correctly orientate the basic region, with respect to the major groove of DNA, and therefore stabilise the resulting ferrocene peptide dimer-DNA complex.

The 48AA ferrocene dimer has been determined as the minimum sequence required to bind DNA specifically. 50AA and 52AA have extremely similar sequences except the latter has a Gly residue beyond the Cys, the site of dimerisation, located at the C-terminus. They are the most helical ferrocene peptide dimers in the presence of CRE DNA as determined by CD, which is again supported by MD simulations performed on model complexes with CRE DNA. However they are over 100 % folded based on our CD calculations. This is due to a protein-induced DNA bending of the CRE DNA, which has been observed in the CD spectra, and which must also contribute to the signal at 222 nm used to calculate the percentage folded.

The electrochemistry of the ferrocene in peptide the dimers was studied using two different sets of experimental setup. In the set with a small volume sample holder, the DPV titration of 52AA Fc yielded the peak currents which were proportional to peptide concentration; however the DPV cannot provide much information about the electrochemistry of the ferrocene in peptide dimers. Whereas CV includes two reversible scans and thus provide more information than the DPV. Using the experimental set-up with all electrodes in the sample solution in a small volume cell, all the ferrocene peptide dimers were studied by CV which shows it is a one electron redox process. The electrochemistry of the peptide dimers is fully reversible under these conditions in the CV.

Chapter 5 DNA binding of Ferrocene peptide dimers

5.1 Introduction

In view of the attractive electrochemical properties, the ability to substitute both Cp rings to generate a symmetrical dimerisation unit, and the effectively (low energy barrier) free rotation of the two Cp rings providing the basic domains with no constraints in order to position themselves for effective DNA binding. We aimed to dimerise GCN4 using ferrocene. Though we are interested in using this as an artificial dimerisation domain in other biological systems, the well-studied GCN4 transcription factor will allow us to evaluate its effectiveness. Studies reported here investigate DNA binding and sequence selectively. How the electrochemistry of the ferrocene probe changes on binding to target and non-specific DNA is investigated and reported here.

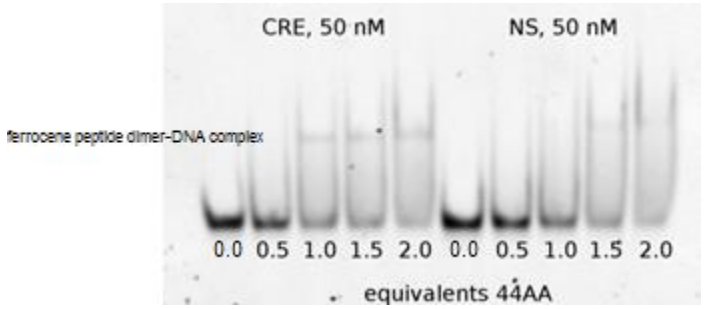
5.2 Gel electrophoresis of ferrocene peptide dimers with CRE and NS DNA

In collaboration with Dr Aimee Gamble, gel electrophoresis experiments were conducted on 50 nM CRE and NS DNA at room temperature with increasing concentrations of ferrocene peptide dimers. Though the experiments were conducted by Dr Gamble in the School of Bioscience, the results will be compared with the CD results presented in the previous section.

In the gel electrophoresis experiment with the 44AA ferrocene dimer, the peptide does not form a new complex with NS DNA. A small amount of 44AA appears to form a new complex with CRE when at least 1 equivalent is added. This is observed as a faint band towards the top of the gel which migrates more slowly due to complex formation, see Figure 1. However this complex does not cause an increase in helicity

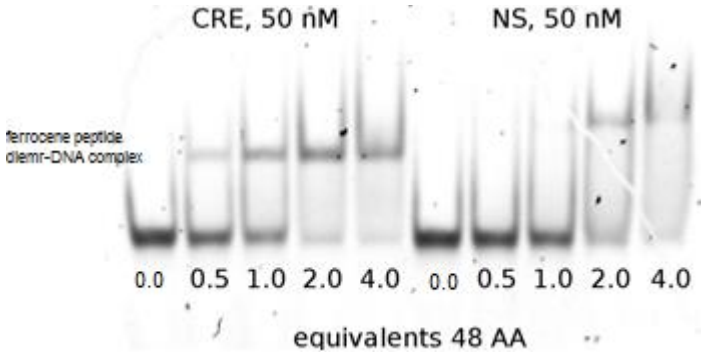
in CD study, so 44AA may interaction with CRE in a non-specific manner which does not involve an α-helical recognition domain.

Figure 1 Gel mobility shift assay of CRE and NS DNA in the presence of increasing equivalents of the 44AA ferrocene dimer.



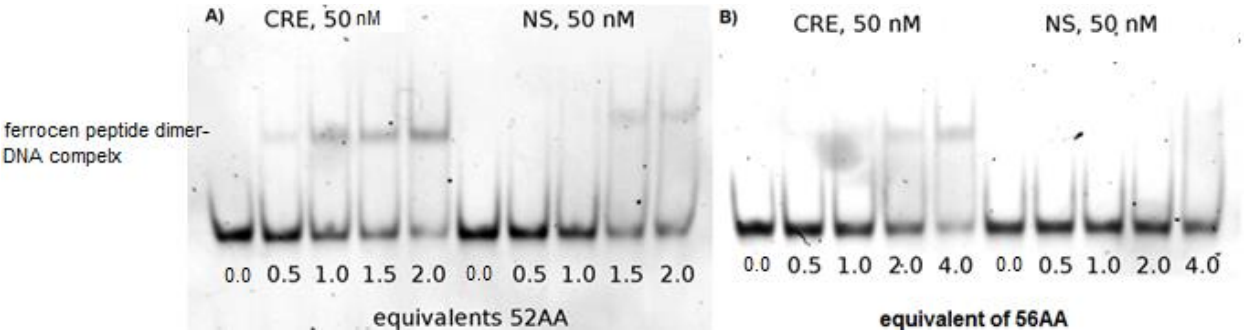
A similar gel electrophoresis study of the 48AA ferrocene dimer, shows the formation of a complex with the CRE DNA to a greater extent than with the 44AA ferrocene dimer. A small amount of the 48AA ferrocene dimer forms a complex with the NS DNA, however this requires significantly more 48AA ferrocene dimer than with the CRE DNA, see Figure 2. Therefore, though same non-specific DNA binding is detected at higher equivalence, the 48AA ferrocene dimer binds sequence selectively at lower equivalents. Furthermore, the formation of a 48AA ferrocene dimer-NS DNA complex is not observed in the CD of the 48AA ferrocene dimer with NS DNA, suggests that the dimer interaction with NS-DNA is non-specific and does not involve the formation of an α-helical recognition domain.

Figure 2 Gel mobility shift assay of CRE and NS DNA in the presence of increasing equivalents of the 48AA ferrocene dimer.



In the gel electrophoresis experiments conducted with the 52AA and 56AA ferrocene dimers and CRE DNA and NS DNA, the results were similar to those obtained for the 48AA ferrocene dimer, consistent with their similar behaviour by CD, see Figure 3.

Figure 3 Gel mobility shift assay of CRE and NS DNA in the presence of increasing equivalents of A) the 52AA ferrocene dimer; B) the 56AA ferrocene dimer.



5.3 Cyclic Voltammetry (CV) of ferrocene peptide dimers with DNA

Having established that some of the ferrocene peptide dimers are able to sequence specifically bind to DNA containing the CRE site with a defined peptide fold, it remained for us to evaluate if the electrochemistry of the ferrocene dimerisation domain was sensitive to these binding events. Therefore, cyclic voltammetry (CV) experiments were conducted in the absence and presence of DNA containing either the CRE or NS sequence.

These experiments were carried out in experimental setup with all working, counter and reference electrodes submerged directly in the sample solution rather than the low volume setup describe in Chapter 3, as this kit has been proved to give valid data for small ferrocene dimer molecules, see Chapter 4. All these CV experiments were independently repeated for 100 μM ferrocene peptide dimers in the presence and absence of one equivalent NS or CRE DNA in triplicate in order to validate the data.

Our ferrocene peptide dimers are studied in solution for DNA detection, and thus the electron transfer occurs through the electrolyte. The basic domains of the peptide, which interact with the CRE DNA site, as well as the CRE DNA, are not expected to play a direct role in the electron transfer. Our hypothesis is that the DNA will cause a change in the ferrocene group redox behaviour, as a result of through space electrostatic interactions. Also the conformational changes of ferrocene peptide dimers on complexation with CRE DNA will cause a conformational change of the ferrocene which may alter its redox behaviours, and would change the solution diffusion behaviour of the redox probe.

Except for 50AA, all the CV of all the ferrocene peptide dimers experiences some similar change in the presence of CRE and NS DNA. In the presence of CRE and NS DNA, all the peptide dimers have a $|i_{pa}/i_{pc}|>1$ and ΔE_p of less than 59 mV, which indicates that none of these ferrocene peptide dimers, have a fully reversible redox couple in the presence of DNA, see Table 1 and Figure 4. In the presence of either CRE or NS DNA, there is cathodic shift in E_{pa} and $E_{1/2}$ as well as an increase in i_{pa} for all the ferrocene peptide dimers (see Table 1 and Figure 4), due to the through space electrostatic interactions between the ferrocene group and the negatively charged DNA, as well these electrostatic interactions making it easier to oxidise. We measured the distance between the nearest DNA phosphate backbone and the ferrocene groups in the simulation models of the ferrocene peptide dimers-CRE DNA complex at 10 ns, see table 2. The ferrocene peptide dimer with a shorter distance from the DNA phosphate backbone has a smaller redox potential ($E_{1/2}$) and thus it is easier to be oxidised.

For the CVs of the 50AA ferrocene dimer in the presence of CRE DNA, the $|i_{pa}/i_{pc}|=1.1$ while ΔE_p is 59 mV, so 50AA is reversible in the presence of CRE. However no MD simulations were performed for the 50AA ferrocene dimer-CRE DNA complex, and therefore the reversible redox behavior of the 50AA ferrocene dimer-CRE DNA complex is unclear.

The CV of 50AA and 52AA in the presence of either CRE or NS DNA, are very different, and the only difference in their structure is the Gly located beyond C-terminus outside the dimerisation unit. The unstructured terminal Gly may hinder the ferrocene in the 52AA-DNA complex, possibly making the complex more difficult to be oxidised and resulting in a more positive E_{pa}. This has also been found for the

56AA ferrocene dimer which contains Gly residues towards the C-terminus, and there is only a 9 to 14 mV decrease in the redox potential of the ferrocene group in the 56AA ferrocene dimer complexed with CRE or NS DNA.

Figure 4 Cyclic voltammetry of 100 µM ferrocene peptide dimers recorded in the absence and presence 100 µM CRE and NS DNA at 20 mV/s, A) 44AA, B) 48AA, C)50AA, D)52AA, E)56AA.

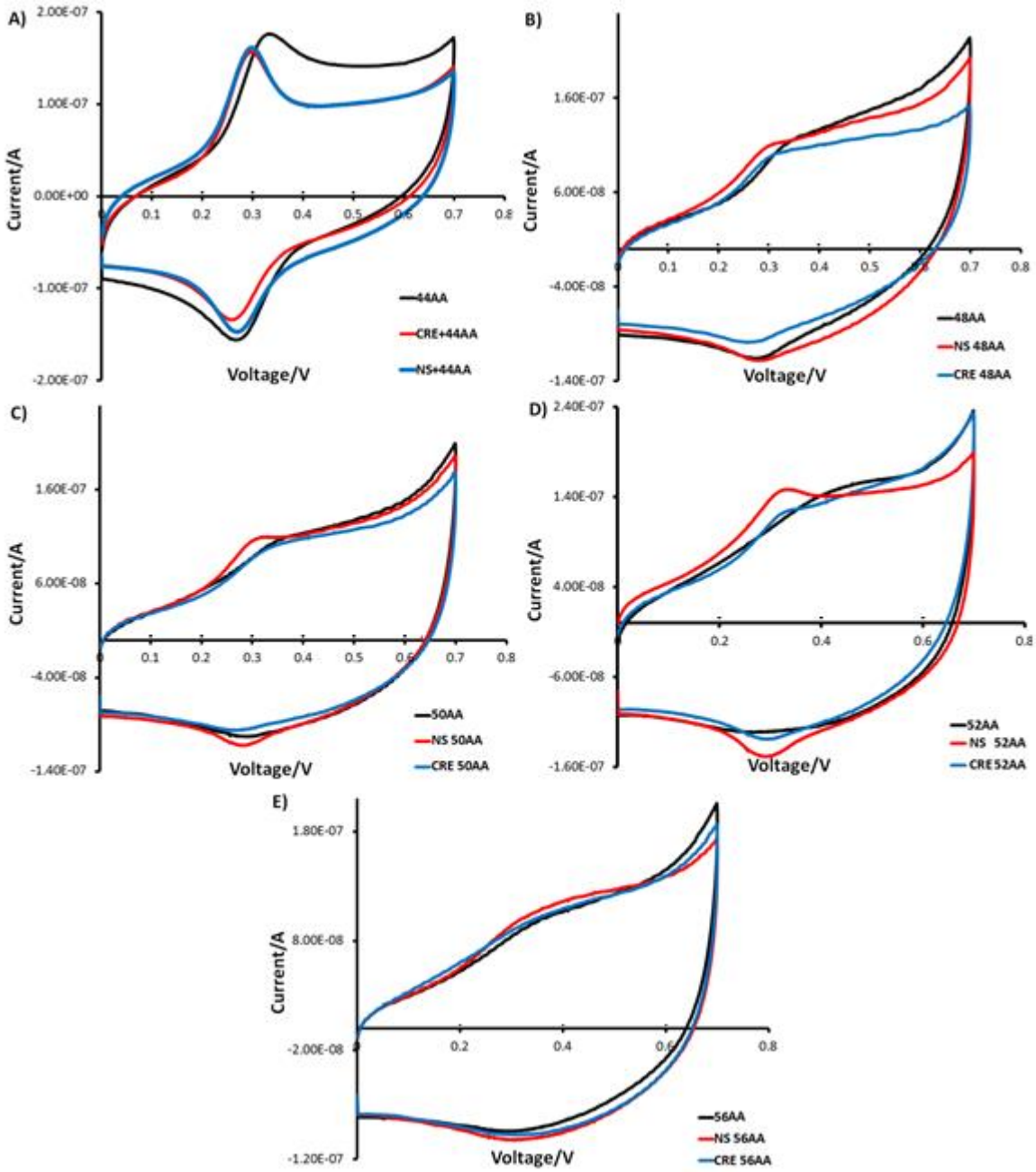


Table 1 The data from the CV of ferrocene peptide dimers in the absence and presence of CRE and NS DNA related to Ag/AgCl. $\Delta E_p = E_{pa} - E_{pc}$, $E_{1/2} = (E_{pa} - E_{pc})/2$

	ipa/A	ipc/A	E_{pa}/mV	E_{pc}/mV	$\lvert i_{pa}/i_{pc} \rvert$	ΔE_p/mV	$E_{1/2}$/mV
44AA	$8.70\times10^{-8}\pm3.48\times10^{-9}$	$-8.59\times10^{-8}\pm4.30\times10^{-9}$	329±8	271±7	1.01	58	300
NS+44AA	$9.97\times10^{-8}\pm4.49\times10^{-9}$	$-5.40\times10^{-8}\pm-1.62\times10^{-9}$	300±6	270±6	1.85	30	285
CRE+44AA	$1.05\times10^{-7}\pm3.14\times10^{-9}$	$-7.80\times10^{-8}\pm-2.73\times10^{-9}$	307±5	252±5	1.34	55	280
48AA	$1.74\times10^{-8}\pm7.48\times10^{-9}$	$-1.73\times10^{-8}\pm-6.40\times10^{-9}$	320±5	262±6	1.01	58	291
NS+48AA	$2.49\times10^{-8}\pm8.47\times10^{-9}$	$-2.26\times10^{-8}\pm-7.46\times10^{-10}$	306±6	279±5	1.1	27	292
CRE+48AA	$3.56\times10^{-8}\pm1.14\times10^{-9}$	$-3.25\times10^{-8}\pm-1.04\times10^{-9}$	310±4	257±5	1.1	53	284
50AA	$1.97\times10^{-8}\pm8.47\times10^{-9}$	$-1.95\times10^{-8}\pm-6.65\times10^{-10}$	358±6	301±4	1.01	57	330
NS+50AA	$3.32\times10^{-8}\pm1.36\times10^{-8}$	$-2.20\times10^{-8}\pm-7.59\times10^{-10}$	306±4	272±4	1.51	34	289
CRE+50AA	$2.16\times10^{-8}\pm7.99\times10^{-9}$	$-1.90\times10^{-8}\pm-7.81\times10^{-10}$	315±5	256±5	1.137	59	286
52AA	$1.90\times10^{-8}\pm9.69\times10^{-10}$	$-1.87\times10^{-8}\pm-3.93\times10^{-10}$	357±8	297±6	1.02	60	327
NS+52AA	$2.55\times10^{-8}\pm1.10\times10^{-9}$	$-2.28\times10^{-8}\pm-7.32\times10^{-10}$	320±9	274±4	1.12	46	297
CRE+52AA	$3.96\times10^{-8}\pm1.27\times10^{-9}$	$-2.79\times10^{-8}\pm-6.44\times10^{-10}$	341±9	290±5	1.42	51	316
56AA	$1.29\times10^{-8}\pm3.74\times10^{-10}$	$-1.20\times10^{-8}\pm-1.74\times10^{-10}$	357±9	300±4	1.08	57	329
NS+56AA	$2.53\times10^{-8}\pm5.82\times10^{-10}$	$-1.56\times10^{-8}\pm-5.79\times10^{-10}$	328±6	300±5	1.62	28	314
CRE+56AA	$2.93\times10^{-8}\pm1.32\times10^{-9}$	$-1.46\times10^{-8}\pm-6.29\times10^{-10}$	345±8	293±7	2.01	52	319

Table 2 The distance between ferrocene in each ferrocene peptide dimer and the nearest phosphate backbones on each CRE DNA in MD simulation models, and the redox potential of the ferrocene peptide dimer in the presence of CRE DNA.

MD model	distance to phosphate backbones/Å	$E_{1/2}$/mV
44AA+CRE	8.8	279.5
48AA+CRE	10.5	283.5
52AA+CRE	12.3	315.5
56AA+CRE	16.3	319

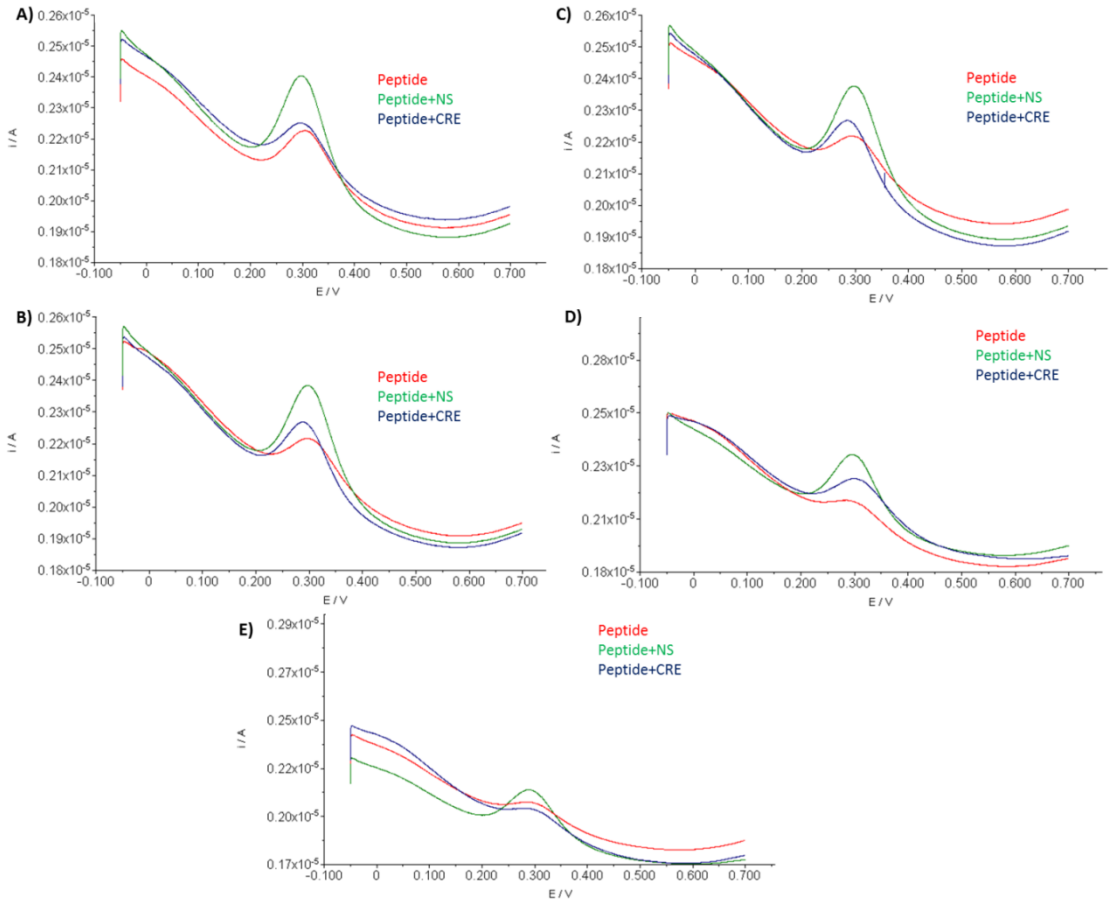
5.4 Differential Pulse Voltammetry (DPV) of ferrocene peptide dimers with DNA

The differential pulse voltammetry (DPV) experiments of ferrocene peptide dimers in the absence and presence of CRE and NS DNA were all conducted in the same experimental setup with the working, counter and reference electrode submerged in the sample solution as in the CV experiments. All the DPV experiments were carried out at the lower concentration of 40 μM, and the ferrocene peptide dimers were added to the DNA solutions to avoid precipitation rather than the reverse. The potential was scanned from -0.05 V to 0.7 V in the DPV, so only the oxidation character of the ferrocene peptide dimers in the absence and presence of DNA was studied, see Table 3 and Figure 5. In the DPV of the ferrocene peptide dimers in the absence of DNA, the dimers with longer peptide strands have smaller peak currents, which is in agreement with the CV. This is because of the steric hindrance introduced by the peptide strands, and similar observation have previously been reported by Jenkins and co-workers.[140] In their study, the presence of a DNA strand attached to a ferrocene shields the ferrocene resulting in a less intense current in the DPV spectrum.[140]

Table 3 The data from the DPV of ferrocene peptide dimers in the absence and presence of CRE and NS DNA related to Ag/AgCl

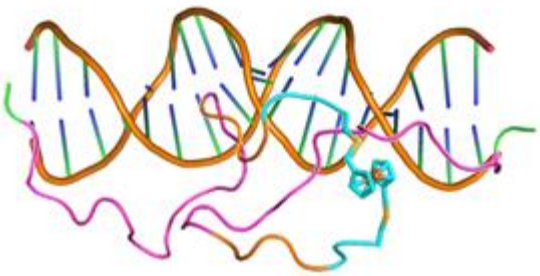
Sample	E_p/mV	I_p/A
44AA	313	1.60×10^{-7}
44AA+NS	304	3.31×10^{-7}
44AA+CRE	305	1.55×10^{-7}
48AA	310	1.25×10^{-7}
48AA+NS	301	3.04×10^{-7}
48AA+CRE	293	1.95×10^{-7}
50AA	307	1.12×10^{-7}
50AA+NS	302	2.95×10^{-7}
50AA+CRE	291	1.91×10^{-7}
52AA	304	9.60×10^{-8}
52AA+NS	300	2.82×10^{-7}
52AA+CRE	303	1.71×10^{-7}
56AA	304	7.29×10^{-8}
56AA+NS	296	1.37×10^{-7}
56AA+CRE	303	8.70×10^{-8}

Figure 5 The DPV of ferrocene peptide dimers in the presence and absence of CRE and NS DNA. A): 44AA ferrocene dimer, B): 48AA ferrocene dimer, C): 50AA ferrocene dimer, D): 52AA ferrocene dimer, E): 56AA ferrocene dimer.



In the DPV study of the 44AA ferrocene peptide dimer, the presence of CRE and NS DNA reduces peak potential, whilst the presence of NS DNA increases the peak current but the presence of CRE DNA does not change the current significantly. The gel electrophoresis study did not indicate the formation of a 44AA ferrocene dimer-CRE DNA complex, while the CD proves that no helix is formed in this complex. Thus the 44AA ferrocene dimer is not structured in the presence of either CRE and NS DNA and binding is non-specific, and in this conformation, the ferrocene is easier to oxidise. In this conformation, the 44AA ferrocene peptide dimer forms random structure on NS DNA, the ferrocene group can get close to the negative charged DNA and thus cause a change in the electrochemical behaviour of the ferrocene group, see Figure 6. The ferrocene group is more exposed to the counter ions in this conformation than its random structure in the absence of DNA.

Figure 6 Cartoon representation of proposed binding mode between our designed ferrocene peptide dimer and NS DNA in a random or non-specific fashion.



In the DPV of the 48AA, 50AA, 52AA and 56AA ferrocene dimers in the presence of NS DNA, the peak current increases. However the CD studies indicated no formation of an α-helix and therefore no binding to NS DNA. These findings were supported by the gel electrophoresis studies. However these did suggest an element of non-specific binding to NS DNA as indicated by same smearing of the DNA, see Figure 6.

Thus the non-helical conformation of the ferrocene peptide dimers on NS DNA expose the ferrocene unit more to the electrolyte than peptide dimers' random structure in the absence of DNA and the dimeric α-helical conformation of peptide dimers in the presence of CRE DNA. The MD simulations of the ferrocene peptide dimers indicates that the ferrocene unit is held between the two peptide strands and is not in full contact with CRE DNA, see Figure 2 in Chapter 4.

The presence of CRE DNA increases the peak currents and decreases the peak potentials of the 48AA, 50AA 52AA and 56AA ferrocene dimer in DPV to make the ferrocene unit easier to be oxidised due to the conformation of ferrocene peptide dimer-CRE DNA exposes the ferrocene unit more to the electrolyte than the ferrocene peptide random conformation in the absence of DNA.

5.5 Conclusions

The ferrocene peptide dimers have been investigated by CD, gel electrophoresis, CV and DPV in the presence of CRE and NS DNA.

The ferrocene peptide dimers similarly display sequence specific binding to DNA, this time based on both CD and gel electrophoresis studies, except for the short 44AA ferrocene peptide dimer, the sequence of which is again too short for recognition of target DNA while its hinge domain sequence is too short to maintain the spatial relationship between the DNA binding basic domain and the ferrocene dimerisation unit. These findings are supported by MD simulation.

The ferrocene linker domains, have been well tolerated, and still allow the peptides to take the advantage of the sequence specific DNA recognition of the native GCN4.

All the ferrocene peptide dimers display fully reversible redox electrochemistry in the absence of DNA based on CV, and thus the peptide does not seem to interfere with the redox advantage of the ferrocene probe. In the presence of CRE or NS DNA, $E_{1/2}$ decreases due to the through space electrostatic interactions between the ferrocene group and the negatively charged DNA. This is also in agreement with the DPV experiments. In the presence of NS DNA, the ferrocene group can get much closer to the DNA due to non-specific binding in a random conformation, and leads to a more significant change in the electrochemistry of the ferrocene group. There may be a relationship between the ferrocene group $E_{1/2}$ and the distance between the ferrocene reporter group and DNA phosphate backbone.

The current for the ferrocene peptide dimer increases in the presence of DNA in DPV, and is always higher in the presence of NS DNA. The random conformation of the ferrocene dimer bound to NS DNA may expose the ferrocene unit more to the counter ions for electron transfer than its random structure in the absence of DNA, or its well folded dimeric binding conformation in the ferrocene dimer-CRE DNA complex, which is supported by the MD models of ferrocene dimer-CRE DNA complex.

For the first time we have designed an electrochemical probe combining the sequence specific DNA binding of GCN4 bZip, and the reversible redox electrochemical behaviour of ferrocene. Though DNA binding is sequence specific, the electrochemical output is not fully understood and cannot be used to identify the nature of the DNA present.

Chapter 6 Conclusions and future work

6.1 Conclusions

A strategy was developed for the dimerisation of two peptides based on GCN4 with a ferrocene linker. This involves: the preparation of a small molecule ferrocene linker; its coupling with the single amino acid, Cys, and the tripeptide glutathione (GSH); synthesis of five peptides based on GCN4 (22AA, 24AA, 25AA, 26AA and 28AA) and their dimerisation with ferrocene. Molecular dynamic simulations were performed on models of four of these ferrocene peptide dimers complexed to target DNA to aid in the design process. Circular dichroism experiments of the ferrocene peptide dimers and peptide monomers in the presence and absence of CRE target site containing DNA and non-specific DNA, were performed in order to evaluate sequence specific DNA binding. Electrochemical studies of the ferrocene peptide dimers in the presence and absence of CRE and non-specific DNA were performed by cyclic voltammetry (CV) and differential pulse voltammetry (DPV).

We attempted to dimerise peptides with a ferrocene linker molecule using two different strategies. Strategy A involved the formation of a peptide bond between the carboxylic acid, located at the C-terminus of an amino acid side chain protected peptide, and an amine of the ferrocene linker molecule. An efficient method for the solid-phase peptide synthesis (SPPS) of a fully side chain protected peptide was developed by the synthesis of a short octapeptide using three different methods. A side chain protected 29 amino acid peptide, designed based on GCN4, was prepared using the best method. However, the product displayed very poor solubility in water and organic solvents, and could not be purified or characterised without fully deprotecting all of the amino acid side chains. Attempts were made to couple this peptide using a commercially available stilbene linker through amine groups.

However, only small amounts of the mono substituted peptide could be detected by MALDI MS. Due to the low solubility of the amino acid side chain protected peptide, this strategy was abandoned. However an unstable 1,1'-diamino ferrocene linker was being prepared for strategy A prior its being abandoned. In strategy B, coupling to the peptide was proposed to occur between the thiol group of a single Cys in the sequence, and a ferrocene linker. A number of ferrocene linkers were initially proposed and synthesised for strategy B, but finally 1,1'-bis (hydroxymethylene) ferrocene was prepared from ferrocene through two reactions. It was first coupled with cysteine and a tripeptide glutathione, and the products fully characterised. These ferrocene peptide dimer model compounds share an absorbance at 255 nm with the same extinction coefficient in their UV-visible spectra which was subsequently used for the quantification of the ferrocene peptide dimers in solution. CV experiments of 1,1'-S-ferrocenyldimethyl-L-glutathione and 1,1'-bis(hydroxymethylene)ferrocene were performed using two different experimental set-ups in order to trial the use of a small volume sample holder, important when working with small quantities of product as was the case here. Unfortunately 1,1'-S-ferrocenyldimethyl-L-glutathione was not found to be electrochemically reversible in the CV experiment using the small sample holder, but instead, the experimental set-up with all three electrodes submerged directly in the sample solution had to be employed in further studies.

We designed five peptides based on the basic and hinge domain of GCN4. Different length peptides, with varying number of conserved residues from native GCN4, were investigated for binding to DNA as monomers. These peptides generally contain the basic region residues, the complete hinge area and an additional short spacing

sequence with a Cys for coupling to the ferrocene linker molecule. All five peptides were coupled to 1,1'-bis (hydroxymethylene) ferrocene to yield the corresponding ferrocene peptide dimers.

The ferrocene peptide dimers were evaluated electrochemically using the two different experimental set-ups described earlier. A DPV titration of the 52AA ferrocene peptide dimer was successfully performed with the small sample holder; however, DPV cannot provide information about the reversibility of the redox chemistry of the ferrocene group. CV experiments of the ferrocene peptide dimers at various scan rates demonstrated that the experimental set up without the small volume sample holder is required. All the ferrocene peptide dimers were studied by CV which shows they undergo a one electron redox process and are fully reversible.

The GCN4 basic domain is unstructured in the absence of DNA containing a target site, and this therefore provides us with a method by which to test the peptide dimers' binding to different DNA sequences by monitoring their folding. Except for the shortest 22AA, all the peptide monomers bind sequence selectively to CRE DNA. The same can be said for the peptide dimers. The shortest 44AA does not bind to DNA while all the longer ferrocene peptide dimers are more helical in the presence of CRE DNA, but are largely unchanged in the presence of NS DNA. The 52AA and 50AA ferrocene peptide dimers have the highest α-helicity in the presence of CRE DNA of all the dimers. However, this is consistent with over 100% folded, and thus there must be another conformation in the complex contributing to the CD signal. Bending of CRE DNA has also been reported in the literature for other DNA binding peptide dimers based on GCN4 and was indicated by the MD simulations of the ferrocene peptide dimer-CRE DNA complexes.

The gel electrophoresis studies investigating binding between the ferrocene peptide dimers and CRE or NS DNA are in agreement with the results obtained from the CD experiments. The shortest 44AA ferrocene peptide dimer does not form a stable complex with CRE DNA, but the remaining ferrocene peptide dimers do form stable complexes with CRE DNA, and these are capable of sequence specific DNA binding.

CV experiments were carried out for all the ferrocene peptide dimers in the presence of one equivalent of CRE or NS DNA, in an effort to study the ferrocene group's redox behavior in the peptide-DNA complex. Only the 50AA ferrocene peptide dimer is electrochemically reversible in the presence of CRE DNA, all the remaining ferrocene peptide dimers are not. All the dimers' E_{pa} and $E_{1/2}$ shift towards negative potential, while there is no trend for the change of E_{pc}, accompanied by an increase in i_{pa}, indicating that the ferrocene groups are easier to oxidise in the presence of DNA due to their proximity to the negatively charged DNA. However, no notable distinction could be made between CRE and NS DNA. There may be a relationship between the ferrocene group $E_{1/2}$ and the distance between the ferrocene group and the DNA phosphate backbone, which is supported by the MD simulations of models of the ferrocene peptide dimer-CRE DNA complex.

In the DPV of the 44AA, 48AA, 50AA, 52AA and 56AA ferrocene peptide dimers recorded in the presence of NS DNA, the peak current increased more than in the presence of CRE DNA. Presumably the ferrocene unit is more exposed to the electrolyte in the presence of NS DNA than in either the peptide dimers' random structure in the absence of DNA, or the α-helical conformation of ferrocene peptide dimer bound to CRE DNA. The latter can be explained by the MD simulations of the ferrocene peptide dimer-CRE DNA complexes, which indicates that the ferrocene

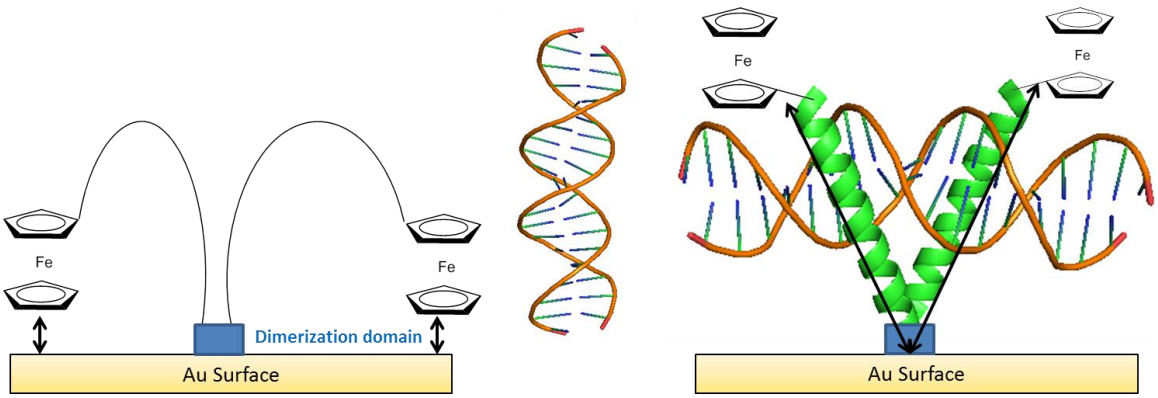
unit is held between two peptide strands and is not fully exposed to the solvent and electrolyte. Though no simulations have been performed on the peptide dimer bound to NS DNA, clearly there is non-specific binding occurring, which is resulting in an increase in the DPV current. The ferrocene peptide dimers display a bigger increase in DPV current in the presence of NS DNA rather than DNA containing the target site, and thus what we have developed is more like a 'negative' electrochemical sensor.

6.2 Future work

6.2.1 Surface bound ferrocene tagged peptide dimer

All the ferrocene peptide dimers reported here were studied in solution in an effort to sense the DNA in our electrochemistry studies, and thus electron transfer occurs through the counter ions in the electrolyte to the ferrocene group attached to the peptide. The peptide sequence (primarily basic domain) and the DNA do not play a direct role in the electron transfer process. In order to involve the peptide strand and the DNA in the electron transfer process, the redox active ferrocene unit can be attached to the N-terminus of the DNA binding peptide dimer, which is attached to an electrode surface, see Figure 1.

Figure 1 Diagram illustrate the conformational change of the electrode surface bound ferrocene peptide dimers; arrows shows the electron transfer occurrence.



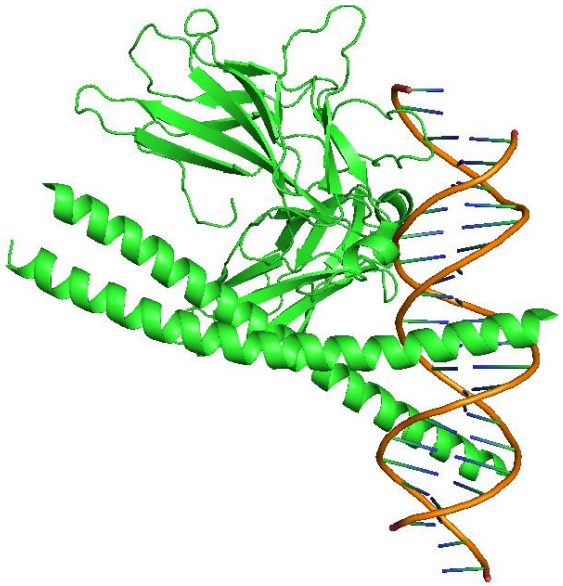
In the absence of DNA, the peptide strands will be random coils in solution, and the redox ferrocene probe will be able to be located in close proximity to the electrode surface for electron transfer. The rate of electron transfer decay across a hydrogen bond is equal to that across two covalent bonds.[141] Therefore, the hydrogen-bonded network may act as a molecular 'shortcut' for electron transfer. When peptide ferrocene dimers bind to the target DNA major groove site, the peptide strand will become helical and produce both an intra and inter hydrogen bonding network with the target DNA, see Figure 1. Therefore the redox active ferrocene group is still accessible for electron transfer, but located further away from the electrode and with more electron transfer decay than that in the absence of DNA. This design may provide an approach for DNA sensing.

6.2.2 Ferrocene dimerised c-Fos/c-Jun heterodimer

The transcription factor c-Fos and c-Jun heterodimer bZip can bind to DNA AP-1 sites. Jun can form a homodimer and a heterodimer with Fos, while Fos cannot form homodimer. The c-Fos and c-Jun heterodimer binds to the AP-1 DNA site 25 times more efficiently than the c-Jun homodimer.[142] The Fos-Jun heterodimer bends DNA towards the major groove, whereas the Jun homodimer bends DNA towards the minor groove to give a distinct DNA-protein complex in topology.[143] The Fos-Jun dimer basic domain consists of residues 139 to 165 of c-Fos, and 254 to 280 of c-Jun, which interacts directly with target DNA.[144] The C-terminal regions form an asymmetric coiled-coil as the dimerisation domain, see Figure 2. Ferrocene dimerisation of the c-Fos and c-Jun basic domains may provide DNA binding

ferrocene peptide dimers similar to our peptide dimers based on GCN4, but would importantly expand the scope to non-symmetrical biomolecular recognition.

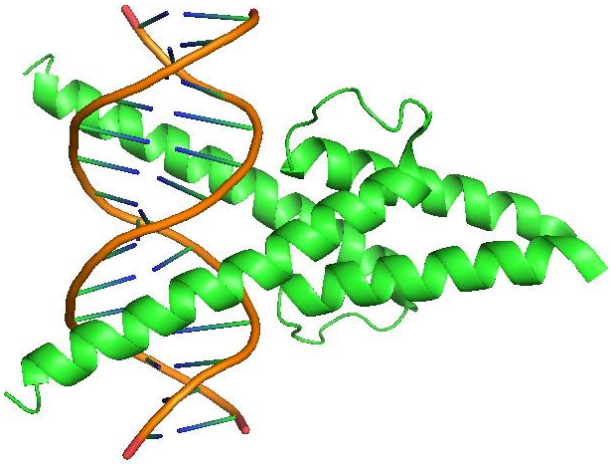
Figure 2 X-ray crystal structure of c-Fos and c-Jun heterodimer bound to DNA (PDB code:1A02).[145]



6.2.3 Ferrocene dimerised helix-loop-helix basic domains

Morii and co-workers prepared four constrained peptide dimers, the sequences of which were based on a helix-loop-helix basic domain of a DNA binding protein MyoD, see Figure 3.[146] The basic domains of MyoD were dimerised with a derivative of 9, 10-dihydrophenanthrene- 9, 10-diol (DHP), (*R,R*)-DHP and (*S,S*)-DHP, with different orientations (N-terminus to N-terminus and C-terminus to C-terminus arrangements). One of the dimers with right handed and C-terminus to C-terminus arrangement, binds specifically to the native MyoD binding site, Figure 11B in Chapter 1.[146]

Figure 3 X-ray crystal structure of MyoD basic helix loop helix domain bound to DNA. (PDB code 1MDY)[147]



The ferrocene replacement for the constrained linker may give a peptide dimer with effectively free rotation of the two Cp rings, providing the basic domains with no constraints in order to position themselves for effective DNA binding. Again this work would expand the scope of these ferrocene peptide dimerisation domains to a different class of DNA binding proteins.

Chapter 7 Experimental Section

7.1 Materials

Ferrocene, aluminium chloride and n-butyllithium were purchased from Acros Organics, acetyl chloride, sodium hypochlorite, 4-dimethylaminopyridine, copper(I) chloride, sodium azide, TFE, anisole, thioanisole, EDT, phenol and acetic anhydride were purchased from Fisher scientific. Bromoacetic acid and phthalimide were purchased from Alfa Aesar. Dimethylmethylene ammonium iodide, tetramethylethylenediamine, 1,1,2,2-tetrabromoethane, copper (I) oxide, ninhydrin and sodium bromohydride were purchased from Sigma Aldrich. Pd 10% WT on active carbon was purchased from Aldrich, TFA, DIPEA, dimethylformamide, HBTU, HOBt, PyBOP, glutathione, resins and all Fmoc protected L-amino acids were purchased from AGTC bioproducts. All the reagents were used as received unless stated otherwise.

7.2.1 Solid-Phase Peptide Synthesis (SPPS) of protected peptide

In manual SPPS, synthesis is performed either at room temperature or sometimes using a microwave CEM discovery unit. The 2-chlorotrityl chloride resin was swollen in DMF 30 min prior the synthesis. Synthesis was performed in the following sequence; wash with DMF (10 mL), add deprotect (10 mL of 20% piperidine in DMF), five washes of 7 mL DMF, add fmoc-protected amino acid (2M in DMF, 5 equivalents), add activator (0.5M HBTU in DMF, 5 equivalents), add activator base (2M DIEA in NMP, 10 equivalents), followed by coupling for 18 minutes. The only exception was Arginine which was double coupled at room temperature for 80 minutes. Following the deprotection of the last amino acid the peptide was capped

with an acetyl (Ac) group by treatment of the resin bound peptide with 20 % acetic anhydride and 20 % DIEA in DMF for 30 minutes at room temperature. The resin was washed three times with DMF (10 mL), followed by three further washings with DCM (10 mL), and finally three washes with diethyl ether (10 mL), before being left to dry overnight. The protected peptide was cleaved by treatment with 20% TFE and 20% acetic acid in DCM, for 4.5 minutes. The cleavage solution after filtration was reduced to a small volume. Addition of cold diethyl ether (-20 °C) results in precipitation of the peptide. MS and analytical HPLC were recorded for short side-chain-protected peptides. However the 29 amino acid side-chain-protected peptide could not be characterised due to solubility issues. In the end the protected peptide was deprotected by TFA/thioanisole/EDT/anisole 90/5/3/2 (10 mL) for characterisation.

7.2.2 Solid-Phase Peptide synthesis by CEM Liberty 1 peptide synthesiser

Peptides (22-28 amino acids) were synthesised using a microwave reactor CEM liberty One automated Peptide Synthesiser on a 0.25 mmol scale with rink amide MBHA resin following standard Fmoc solid-phase peptide synthesis (SPSS) protocols. The synthesis was carried out in a fritted polymer reaction vessel. The resin bubbled by nitrogen gas in DMF (15 mL) 30 min prior to the synthesis. Synthesis was performed in the following sequence; wash with DMF (10 mL, top), add deprotect (10 mL of 20% piperidine in DMF with 0.1M HOBT, 35W, 75°C, 180s), four washes of 10 mL DMF (top, bottom, top, top), add amino acid (2M in DMF, 5 mL), add activator (2 mL, 0.5M HBTU in DMF), add activator base (1 mL, 2M *N*,*N*-

diisopropylethylamine in *N*-methyl-2-pyrrolidone), followed by running the coupling program (25W, 75°C, 300s). For the synthesis of the 25AA, 26AA and 28AA peptide, (25W, 75°C, 300s) except for Cys (0W, 50°C, 120s followed by 25W, 50°C, 240s) and Arg (0W, room temperature, 2400s), all the couplings were performed as standard double programs, in which the amino acid, activator and activator base were added and coupled, drained, and the resin washed before addition of a second portion of amino acid, activator and activator base and coupling (25W, 75°C, 300s). In the synthesis of the remaining peptides (22AA, 24AA), all the coupling were standard single programs, except for Cys (0W, 50°C, 120s followed by 25W, 50°C, 240s) and Arg (0W, room temperature, 2400s). Following the deprotection of the last amino acid the peptide was capped with an acetyl (Ac) group by treatment of the peptide with, 20 % acetic anhydride, 20 % *N*,*N*-diisopropylethylamine in DMF for 30 minutes at room temperature. The resin was washed three times with DMF (10 mL), followed by three further washings with DCM (10 mL), and finally three washes with diethyl ether (10 mL), before being left in a desiccator to dry overnight.

The resin bound peptide was stirred with TFA(90%)/thioanisole(5%)/1,2-ethanedithiol(3%)/anisole(2%) (10 mL) for 4 hours under nitrogen gas. This simultaneously cleaves the peptide from the resin and removes the amino acid side chain protecting groups to yield the C-terminal amide (due to choice of resin). The peptides were separated from the resin by filtration and the volume of the TFA peptide filtrate reduced to about 5 mL. To this was added cold diethyl ether (-20 °C) which resulted in the formation of a white precipitate. This suspension was stored at -20 °C overnight, so as to ensure maximum recovery of product. The precipitate was collected by filtration and dried in a desiccator overnight.

The crude peptide was dissolve in 25 mL 10% acetic acid in water, and then purified by preparative reversed-phase C12-HPLC by increasing the gradient from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes. This was followed by the continuous wash with 20% acetonitrile (0.05% TFA) for a further 20 minutes. The collected fractions were characterised by ESI-MS, those that contained the product were combined and lyophilised to yield the purified peptide as a white solid.

The peptides were characterised by ESI-MS and purity established by analytical C12 HPLC with the same gradient as used earlier. Peptide concentrations were determined by the Ellman's test.

22AA ESI-MS: expected: M=2440.9, *m*/*z* observed: $[M+3H]^{3+}$=814.4, $[M+4H]^{4+}$=611.1, $[M+5H]^{5+}$=489.1, $[M+6H]^{6+}$=407.7, consistent with a deconvoluted M=2440.3.

24AA ESI-MS: expected: M=2767.2, *m*/*z* observed: $[M+3H]^{3+}$=923.2, $[M+4H]^{4+}$=692.6, $[M+5H]^{5+}$=554.3, $[M+6H]^{6+}$=462.1, $[M+7H]^{7+}$=396.2, consistent with a deconvoluted M=2766.5.

25AA MALDI-MS: expected: M=3040.6, *m*/*z* observed: $[M+H]^{+}$=3041.5.

26AA ESI-MS: expected: M=3097.6, *m*/*z* observed: $[M+3H]^{3+}$=1033.3, $[M+4H]^{4+}$=775.2, $[M+5H]^{5+}$=620.4, $[M+6H]^{6+}$=517.3, $[M+7H]^{7+}$=443.4, consistent with a deconvoluted M=3097.0.

28AA ESI-MS: expected: M=3211.8, *m*/*z* observed: $[M+4H]^{4+}$=803.8, $[M+5H]^{5+}$=643.2, $[M+6H]^{6+}$=536.2, consistent with a deconvoluted M=3211.1.

7.2.3 Kaiser test

Potassium cyanide in distilled water solution (0.253 mmol, 1 mL) was diluted with pyridine (49 mL) to make solution A; ninhydrin (1 g) was dissolved in *n*-butanol (20 mL) to make solution B; phenol (40 g) was dissolved in *n*-butanol (20 mL) to make solution C. A few beads of resin were put into a glass sample vial; 2-3 drops of solution A, B and C were added to the resin. The mixture was heated at 110 °C for 5 minutes. The resin with free amine turned blue after 5 minutes, while the resin without free amine remained the same colour. This therefore allows for the determination of whether a coupling or deprotection has successfully occurred.

7.2.4 Ellman's Test of peptide

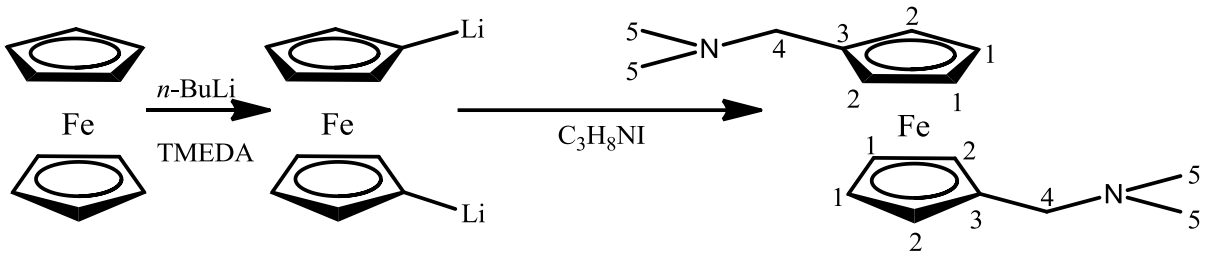
For the Ellman's test of the peptide solution, a blank solution of 3 mL TRIS buffer pH 8 with 0.1 mL 5,5'-dithiobis-(2-nitrobenzoic acid) solution (30 mM) was used as the control blank. This was followed by the addition of 0.01 mL of the peptide stock to the blank solution, and after 15 minutes equilibration the UV-visible spectrum was recorded between 360 nm and 600 nm. The test was repeated in triplicate and the average absorbance at 412 nm was used to calculate the concentration based on the extinction coefficient of 13,600 $M^{-1}cm^{-1}$ for the 2-nitro-5-thiobenzoate product. This allows for the concentrations of readily available thiol groups to be determined.

7.3 Preparation of the low molecular weight disubstituted ferrocene molecules

7.3.1 1,1-Bis(N,N-dimethylaminomethy)ferrocene

1,1-Bis(N,N-dimethylaminomethy)ferrocene preparation was based on the method reported by Glidewell[115] with some modification.

Scheme 1 Reaction scheme for the synthesis of 1,1-bis(N,N-dimethylaminomethy)ferrocene from ferrocene in 2 steps.



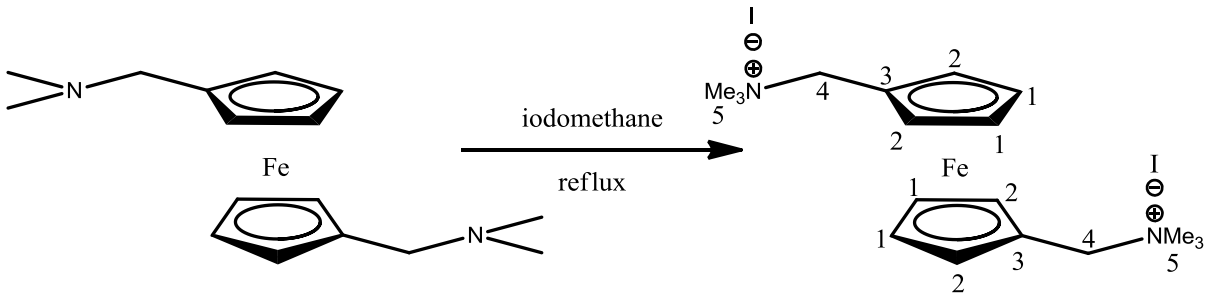
n-Butyllithium (34 mL, 54 mmol) was added dropwise to a stirred solution of ferrocene (4.84 g, 26 mmol) and tetramethylethylenediamine (6.28 g, 54 mmol) in hexane (250 mL) under nitrogen gas for 18 hours, and then diluted with diethyl ether (200 mL). Dimethylmethylidene ammonium iodide (10 g, 54 mmol) was added and the reaction mixture stirred under nitrogen gas for a further 24 hours. Water (50 mL) was added and the reaction mixture was stirred for 20 min. A second portion of water (100 mL) was added. The aqueous layer was washed with diethyl ether (150 mL) twice, the organic layer dried over $MgSO_4$, which was removed by filtration and the solvent removed in *vacuo* to yield the crude product. 1, 1'-Bis (dimethylamino methyl) ferrocene was isolated by flash column chromatography (30 % DCM- 10 % triethylamine -60 % hexane) as a brown solid (1.89 g, 24 %).

δ_H (300 MHz, $CDCl_3$): 2.18 (12H, s, H_5), 3.27 (4H, s, H_4), 4.07 (4H, t, *J* 2 Hz, H_1), 4.10 (4H, t, *J* 2 Hz, H_2).

ESI-MS+: expected M[$C_{16}H_{24}FeN_2$]=300.1, *m*/*z* observed $[M+H]^+$ = 301.1.

7.3.2 1, 1'-Bis(trimethylammonium iodide methyl) ferrocene

Scheme 2 Reaction scheme for the synthesis of 1, 1'-bis(trimethylammonium iodide methyl) ferrocene



The preparation of 1, 1'-bis(trimethylammonium iodide methyl) ferrocene was modified slightly from Forgie's method.[148]

Iodomethane (0.73 g, 0.005 mol) was added dropwise into a stirred solution of 1,1-diacetyl ferrocene (0.5 g, 0.0017 mol) in methanol (5 mL). The mixture was stirred under reflux for 10 min. Precipitation with diethyl ether (150 mL) gave 1, 1'-bis(trimethylammonium iodide methyl) ferrocene as a brown solid (0.95 g, 95 %).

δ_H (300 MHz, D_2O): 2.88 (18H, s, H_5), 4.30 (4H, s, H_4), 4.40 (4H, t, *J* 2 Hz, H_1), 4.90 (4H, t, *J* 2 Hz, H_2).

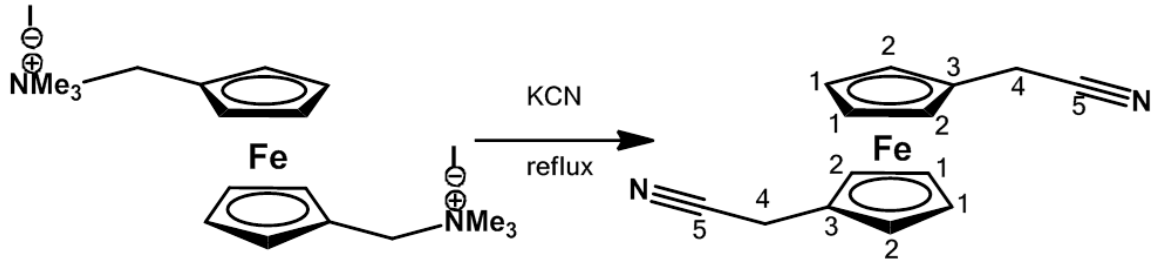
δ_c(75 MHz, D_2O, DEPT): 52.0 (C_3), 66.2 (C_4), 71.8 (C_1), 73.2 (C_2).

ESI-MS+: expected M[$C_{18}H_{30}FeN_2I_2$]= 584.0, *m/z* observed $[M\text{-}I]^+$ = 457.0.

7.3.3 1, 1'-Bis nitrilemethylene ferrocene

1, 1'-Bis nitrilemethylene ferrocene was prepared according to the preparation of cyanomethyl ferrocene by Daniel, but with modified reaction times.[149]

Scheme 3 Synthesis of 1,1'-bis(cyanomethyl) ferrocene by nitrile substitution under reflux.



A mixture of potassium cyanide (1.27 g, 0.0195 mol) and 1, 1'-bis(trimethylammonium iodide methyl) ferrocene (0.95 g, 0.0016 mol) in water (8 mL) was stirred under reflux (120 ℃) for 34 hours. Solvent extraction with diethyl ether, drying over $MgSO_4$, filtration and solvent removal in *vacuo* yielded 1, 1'-bis nitrilemethylene ferrocene (43 mg, 10 %) as a yellow solid, (19 mg, 5%). Half of the residues was dissolved in water (8 mL) and stirred under reflux (120 ℃) for a further 3 days. An additional 1, 1'-bis nitrilemethylene ferrocene was isolated as yellow solid by solvent extraction (diethyl ether), (41 mg 12%). Overall yield (60 mg, 17 %)

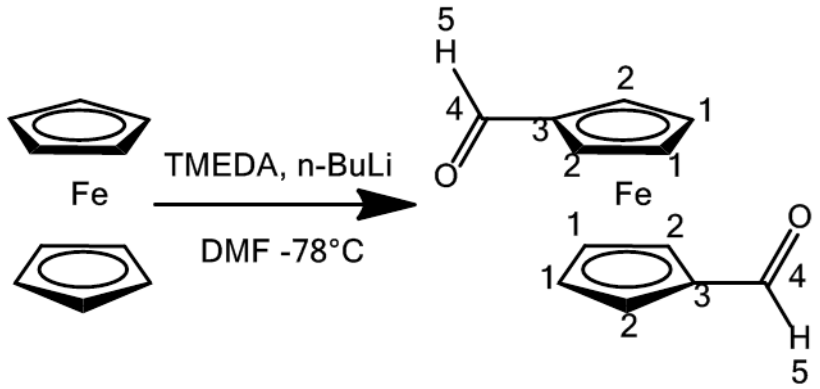
δ_H (300 MHz, $CDCl_3$): 4.10 (4H, t, *J* 2 Hz, H_2), 4.14 (4H, t, *J* 2 Hz, H_1), 4.33 (4H, s, H_4).

ESI-MS+: expected M[$C_{14}H_{12}FeN_2$]=264.0, *m/z* observed [M+H+MeOH]$^+$= 297.1.

7.3.4 1,1'-Ferrocenedicarboxaldehyde

This synthesis is modified version of Connell's report.[117]

Scheme 4 Synthesis of 1,1'-ferrocenedicarboxaldehyde



Tetramethylethylenediamine (8 mL, 0.53 mmol) was added to a stirred solution of ferrocene (5 g, 26 mmol) in dry diethyl ether (60 mL) in one portion under nitrogen gas. The solution was stirred at room temperature for 10 minutes and then cooled to 0℃. n-Butyl lithium (35 mL, 1.6 M, 56 mmol) was added dropwise at 0℃, and the orange solution stirred for 21 hours at room temperature. Anhydrous dimethylformamide (7 mL, 91 mmol) was added dropwise to the mixture at -78℃ and the mixture stirred for 2 hours under a nitrogen gas atmosphere. The mixture was opened to the air, water (50 mL) added and stirred for a further 30 minutes. The aqueous layer was extracted four times with hot ethyl acetate (100 mL), and the combined organic layers were dried over $MgSO_4$. Evaporation to dryness gave the crude product as a dark red solid. Recrystalisation from boiling ethyl acetate gave 1,1'-ferrocenedicarboxaldehyde as red needles(0.9 g, 14 %).

^{1}H NMR (300 MHz, $CDCl_3$): δ= 9.95 (2H, s, H_5), 4.88 (4H, t, *J* 1.5 Hz, H_2), 4.67 (4H, t, *J* 1.5 Hz, H_1).

^{13}C NMR (148 MHz, $CDCl_3$): δ= 193.12 (C_4), 80.67 (C_3), 74.52 (C_2), 71.22 (C_1).

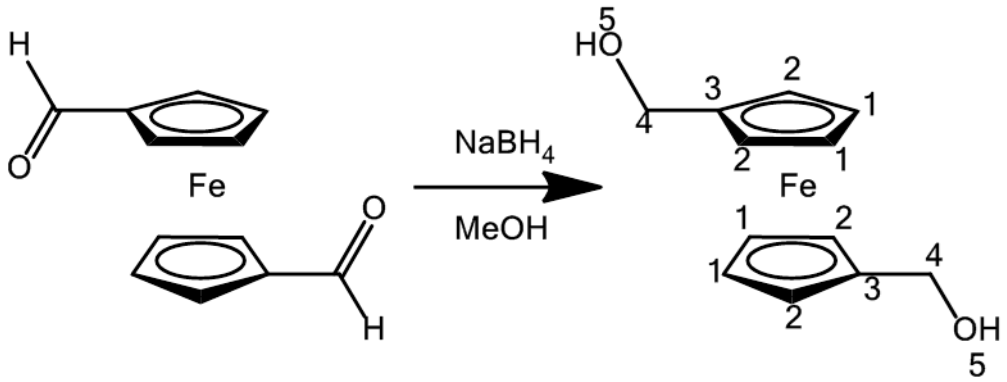
ES-MS+: expected M[$C_{12}H_{10}FeO_2$]= 242.0, *m*/*z* observed $[M+Na]^+$= 265.0.

Melting point: 155-157 ℃; literature melting point: 156 ℃.[150]

7.3.5 1,1'-Bis(hydroxymethylene)ferrocene

This synthesis is modified from Patwa's method.[118]

Scheme 5 synthesis of 1,1'-bis(hydroxymethylene)ferrocene



A solution of 1,1'-ferrocene dicarboxaldehyde (0.31 g, 1.28 mmol) in dry methanol (5 mL) was stirred at 0 ℃ under nitrogen gas for 15 minutes. $NaBH_4$ (0.11 g, 2.82 mmol) was added into the solution in one portion, the mixture was stirred for 2.5 hours under a nitrogen gas atmosphere at room temperature. The reaction mixture was evaporated to dryness, and extracted three times with ethyl acetate (15 mL), the combined organic phases was filtered and evaporated to dryness to yield the crude product as an orange solid. This was purified by flash column chromatography (hexane: diethyl ether=1:1) and precipitation from boiling ethyl acetate on addition of

cold petrol to yield 1,1'-bis(hydroxymethylene) ferrocene as brown needles (0.24 g, 76 %).

^{1}H NMR (300 MHz, $CDCl_3$): δ= 4.40 (s, 4H, H_4), 4.21 (t, 4H, *J* 1.5 Hz, H_2), 4.17 (t, 4H, *J* 1.5 Hz, H_1), 3.21 (s, 2H, H_5).

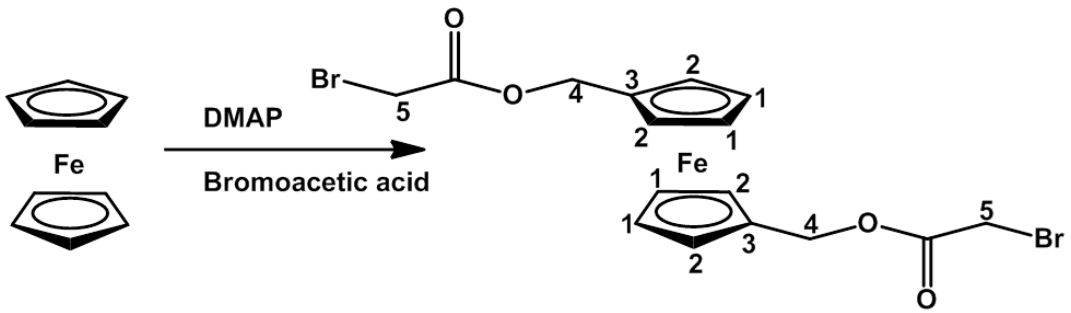
^{13}C NMR (148 MHz, $CDCl_3$): δ= 89.73 (C_3), 68.47 (C_2), 67.50 (C_1), 60.75 (C_4).

ES MS+: expected M[$C_{12}H_{14}FeO_2$]=246.0, *m/z* observed $[M]^+$= 246.0 , $[M+Na]^+$= 269.0.

Melting point: 108-110 ℃; literature melting point= 106-108 ℃.[118]

7.3.6 1,1'-Ferrocenyl bis (methylene 2-bromoacetate)

Scheme 6 Synthesis of 1,1'-ferrocenyl bis (methylene 2- bromoacetate)



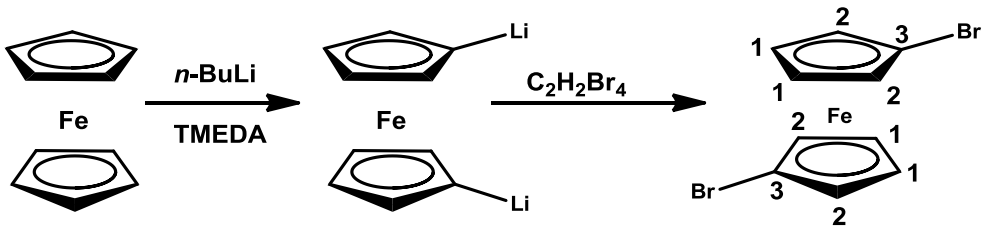
A solution of 4-dimethylaminopyridine (0.5 mg, 0.0041 mmol) in DCM (0.1 mL) was added into a stirred solution of bromoacetic acid (0.17 g, 1.22 mmol) and 1,1'-bis(hydroxymethylene) ferrocene (0.1 g, 0.406 mmol) in DCM (10 mL) solution. The reaction mixture was stirred under reflux (48 ℃) and protected from the light for 48 hours, and then the reaction mixture was evaporated to dryness to yield the crude

product as a brown sticky solid. This was characterised by ^{1}H NMR which indicated bromoacetic acid was present as an impurity. Column chromatography and HPLC were subsequently attempted for the purification of 1,1'-ferrocenyl bis (methylene 2-bromoacetate), however it was found to decomposed after column chromatography and HPLC.

^{1}H NMR (300 MHz, $CDCl_3$): δ= 3.83 (4H, s, H_5), 4.22-4.23 (4H, t, *J* 2 Hz, H_2), 4.29-4.30 (4H, t, *J* 2 Hz, H_1), 4.98 (4H, s, H_4).

7.3.7 1,1'-Dibromo ferrocene

Scheme 7 Synthesis of 1,1'-dibromo ferrocene



The preparation of 1,1'-dibromo ferrocene is based on Shafir's method [24] and the preparation of the intermediate 1,1'-dilithiaferrocene is based on Bishop's method [25].

n-Butyllithium (1.6M in hexane, 34 mL, 54 mmol) was added dropwise into the stirred solution of ferrocene (4.84 g, 26 mmol) and TMEDA (6.28 g, 54 mmol) in degased hexane (200 mL). The mixture was stirred for 20 hours, resulting in the formation of an orange precipitate. Tetrabromoethane (6.65 mL, 19.7 g, 57 mmol) was added dropwise to the fresh dilithium ferrocene in hexane mixture which was stirred in an acetone/dry ice cooling bath (-78 ℃) for 3 hours followed by 14 hours

stirring at room temperature. Precipitation twice gave 1,1'-dibromoferrocene as an orange powder (2.13 g, 24 %).

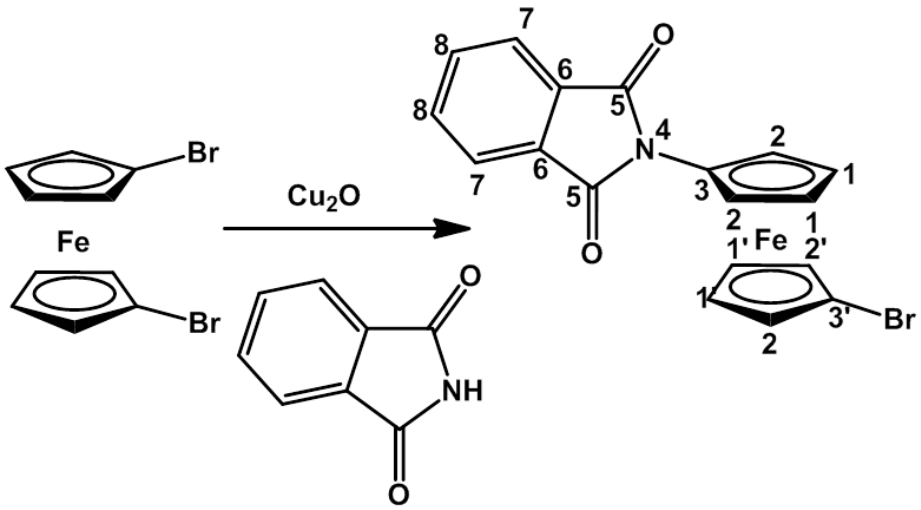
δ_H (300 MHz, $CDCl_3$): 4.16-4.30 (4H, m, H_1), 4.49 (4H, s, H_2).

δ_c(75 MHz, $CDCl_3$, DEPT): 69.9 (C_1), 72.7 (C_2), 78.4 (C_3).

ESI-MS+: expected M[$C_{10}H_8Br_2Fe$]= 343.8, *m*/*z* observed $[M]^+$ = 343.8.

7.3.8 1-Bromo-1'-(1,3-dihydro-1,3-dioxo-2*H*-isoindol-2-yl)-ferrocene[151]

Scheme 8 Synthesis of 1-bromo-1'-(1,3-dihydro-1,3-dioxo-2*H*-isoindol-2-yl)-ferrocene



A mixture of 1,1'-dibromo ferrocene (0.45 g, 1.3 mmol), phthalimide (0.625 g, 4.2 mmol) and copper(I) oxide (0.19 g, 1.3 mmol) in degased pyridine (5 mL), was under nitrogen atmosphere and refluxed (120 °C) for 48 hours. The mixture was evaporated to dryness, triturated with hexane and filtered through a short column of Al_2O_3. The column was washed with hexane (400 mL) until the wash became colourless before the addition of reaction mixture, and then washed with diethyl ether

and the orange ether was evaporated to give the crude product as dark red needle crystals. The crude product was dissolved with hot diethyl ether (5 mL), and recrystallised in an ice bath to give 1-bromo-1'-(1,3-dihydro-1,3-dioxo-2*H*-isoindol-2-yl)-ferrocene as yellow needle crystals (34 mg, 6.4 %).

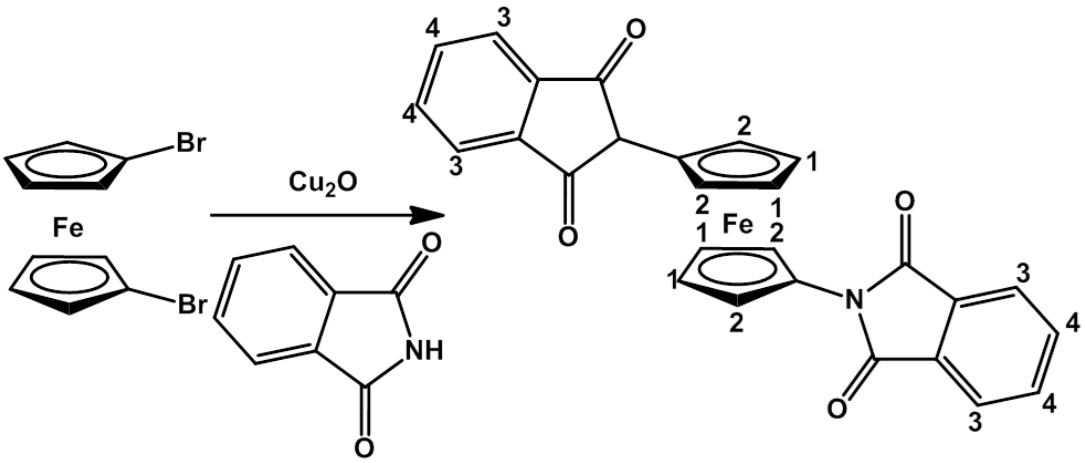
δ_H (300 MHz, $CDCl_3$): 4.15-4.46 (6H, m, H_1, H_2, $H_{1'}$), 5.08-5.21 (2H, m, $H_{2'}$), 7.74-7.77 (2H, m, H_7, H_8).

δ_c(75 MHz, $CDCl_3$): 63.20 (C_3), 63.30 (C_2), 66.92 (C_1), 67.45 ($C_{1'}$), 69.25 ($C_{3'}$), 70.49 ($C_{2'}$), 122.37 (C_8), 131.06 (C_6), 133.33 (C_7). 165.90 (C_5).

ES MS+: expected M[$C_{18}H_{12}FeBrNO_2$]=408.9, *m*/*z* observed $[M+Na-Br]^+$=354.0, $[M+Na+MeOH-Br]^+$=386.0, $[M+Na]^+$=432.0, $[M+Na+MeOH]^+$= 464.0.

7.3.9 1,1'-Bis(1,3-dihydro-1,3-dioxo-2*H*-isoindol-2-yl)-ferrocene[152]

Scheme 9 Synthesis of 1,1'-Bis(1,3-dihydro-1,3-dioxo-2*H*-isoindol-2-yl)-ferrocene



A mixture of 1,1'-dibromo ferrocene (1 g, 2.9 mmol), phthalimide (2.78 g, 19 mmol), and Cu_2O (0.42 g, 5.8 mmol) in degased pyridine (7 mL) was under a nitrogen

atmosphere and refluxed (120 ℃) for 26 hours. The mixture was evaporated to dryness, and then redissolved in DCM, filtered through a hexane washed celite pad and washed with DCM until the wash became colourless. The evaporated filtrate was dissolved in chloroform, washed by water, dried by $MgSO_4$ and evaporated to dryness to give crude 1,1'-bis(1,3-dihydro-1,3-dioxo-2*H*-isoindol-2-yl)-ferrocene as brown solid (0.773 g, 56 %) and this is used for future synthesis as phthalimide and cannot be remove by purification.

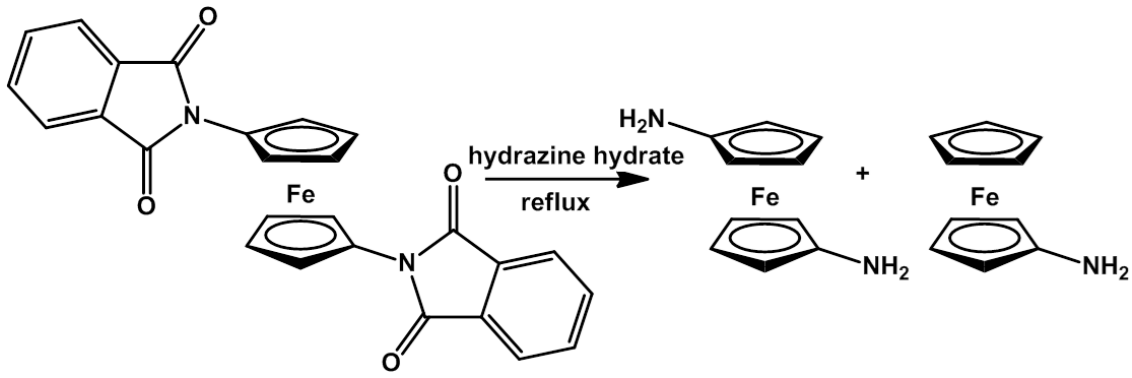
δ_H (300 MHz, $CDCl_3$): 4.23-4.30 (4H, m, H_1), 5.79-5.39 (4H, m, H_2), 7.58-7.61 (8H, m, H_3 and H_4).

ES MS+: expected M[$C_{27}H_{16}FeN_2O_4$]=476.3, *m*/*z* observed $[M+Na]^+$= 499.0.

MALDI MS+: expected M[$C_{27}H_{16}FeN_2O_4$]=476.3, *m*/*z* observed $[M+H]^+$= 477.3.

7.3.10 1,1’-Diamino ferrocene and amino ferrocene

Scheme 10 Synthesis of 1,1’-diamino ferrocene and amino ferrocene



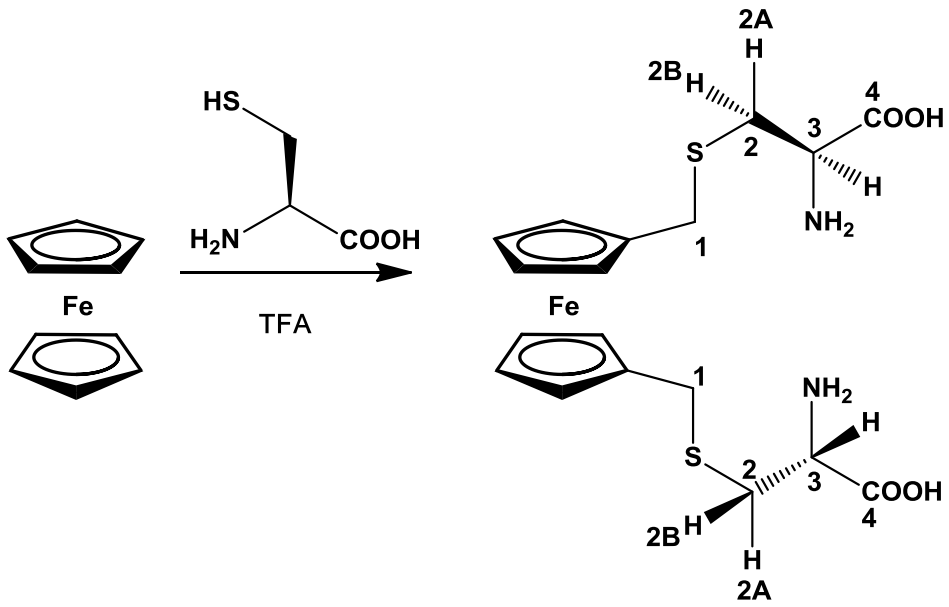
A mixture of hydrazine hydrate (8 mL), 1,1'-bis(1,3-dihydro-1,3-dioxo-2*H*-isoindol-2-yl)-ferrocene (770 mg, 1.6 mmol) in degased ethanol (20 mL) was refluxed for 3 hours under nitrogen gas. The mixture was cooled to room temperature, water (20

mL) was added, extracted with diethyl ether (2×20 mL), dried over $MgSO_4$, which was removed by filtration, solvent removed in vacuo to give the crude product as a yellow powder. The crude product is used for future synthesis directly, because it is unstable and decomposes under nitrogen gas.

ES-MS+: expected M[$C_{10}H_{12}FeN_2$]=216.0, *m*/*z* observed [M]+=216.0, [M-NH_2]=201.0.

7.3.11 1,1'-s-Ferrocenyldimethyl-L-cysteine

Scheme 11 Synthesis of 1,1'-s-ferrocenyldimethyl-L-cysteine



To a solution of 1,1'-bis(hydroxymethylene)ferrocene (0.2 g, 0.9 mmol) in degased acetone (4.4 mL) was added a solution of L-cysteine (0.222 g, 1.8 mmol) in degassed water (3.7 mL), followed by the cautious addition of TFA (0.1 mL). The reaction mixture was stirred under a nitrogen atmosphere for 48 hours. The solution was concentrated under reduced pressure, and the product precipitated by addition of

diethyl-ether. The pale yellow solid was recovered by filtration and washed with diethyl ether and air dried. Yield 95 mg (23%).

^{1}H NMR (300 MHz, DMSO-d_6, 5% TFA): δ= 8.34 (s, 4H, NH_2), 4.11-4.20 (m, 8H, CpH), 4.14-4.12 (m, 2H, H3), 3.60 (s, 4H, H1), 2.93 (dd, *J* 5, 15 Hz, 2H, H2B), 2.85 (dd, J 7, 15 Hz, 2H, H2A).

^{1}H NMR (300 MHz, MeOH-d_4, 5% HCl): δ= 4.32-4.24 (m, 8H, CpH), 4.11 (t, *J* 4 Hz, 2H, H3), 3.62 (s, 4H, H1), 3.14 (dd, *J* 3, 15 Hz, 2H, H2B), 2.95 (dd, *J* 8, 15 Hz, H2A).

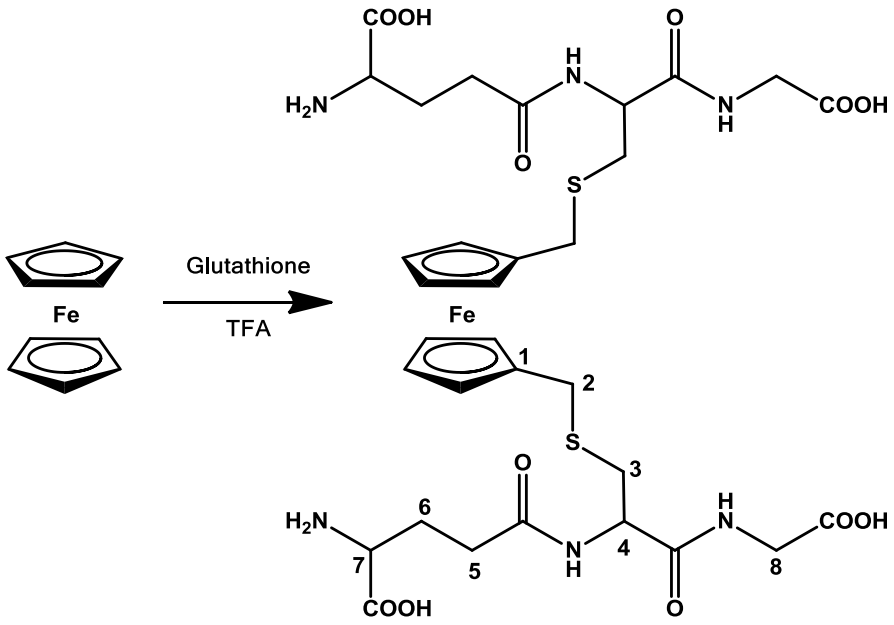
^{13}C NMR (148 MHz, DMSO-d_6, 5% TFA): δ= 169 (C=O), 84 (Cp<u>C</u>-CH_2), 69, 68 (Cp<u>C</u>-H), 51 (C3), 31, 30 (C1, C2).

ES MS: expected M[$C_{18}H_{24}FeN_2O_4$]=452.3, *m/z* observed $[M]^+$= 452.3, $[M\text{-}Cys]^+$= 332.0.

MALDI MS: expected M[$C_{18}H_{24}FeN_2O_4$]=452.3, *m/z* observed 452.3 $[M]^+$, 475.2 $[M+Na]^+$, 491.0 $[M+K]^+$, 332.0 $[M\text{-}Cys]^+$.

7.3.12 1,1'-s-ferrocenyldimethyl-L-glutathione

Scheme 12 Synthesis of 1,1'-s-ferrocenyldimethyl-L-glutathione



To a solution of glutathione (0.56 g, 1.8 mmol) in degassed water (4.2 mL) was added a solution of 1,1'-bis(hydroxymethylene)ferrocene (0.2 g, 0.8 mmol) in degased water-acetone (43%/57% 5 mL) followed by the cautious addition of TFA (0.1 mL). The reaction mixture was stirred for 49.5 hours at room temperature and the solvent removed *in vacuo*, followed by the addition of ethanol (3×8 mL) and its subsequent removal under *vacuo*. The residue was redissolved in 3 mL water, precipitated on addition of ethanol (6 mL) and storage at -20 °C. A yellow sticky solid (0.3 g) was recovered by filtration and lyophilisation overnight yielded a yellow powder. Yield (0. 25 g, 38%).

^{1}H NMR (300 MHz, D_2O): δ= 4.47-4.44 (dd, *J* 3, 8 Hz, 2H, 2×H7), 4.28-4.24 (m, 8H, 8×Cp-H), 3.93 (s, 4H, 4×H8), 3.78 (t, 2H, *J* 6Hz, 2×H4), 3.49 (s, 4H, 4×H2), 3.02-2.96 (dd, *J* 5, 14 Hz, 2H, 4×H3), 2.81-2.77 (dd, *J* 5, 14 Hz, 2H, 4×H3), 2.49-2.44 (td, *J* 3, 8 Hz, 4H, 4×H5), 2.13-2.05 (q, *J* 7 Hz, 4H, 4×H6).

^{13}C NMR (148 MHz, DMSO-d_6): δ= 171.6, 170.9, 170.8, and 170.7 (C=O), 85.2 (C1), 69.3, 69.2, 68.3 and 68.2 (CpC), 52.9 (C7), 52.3 (C4), 41.1(C8), 33.5 (C3), 31.5 (C2), 30.8 (C5), 26.9 (C6).

ES MS: expected M[$C_{32}H_{44}FeN_6O_{12}S_2$]=824.2, *m/z* observed $[M+H]^+$= 825.1, $[M+Na]^+$= 847.0, $[M+2Na-H]^+$= 869.0.

7.4.1 Coupling of the peptide with 1,1'-bis (hydroxymethylene)ferrocene

The synthesis of di substituted ferrocene peptide dimers was performed at 0.9 µmol scale in water (2 mL) with 0.3 µmole 1, 1'-bis (hydroxymethylene) ferrocene in acetone (1 mL). All the solvents were degased by nitrogen gas bubbling for 30 minutes. TFA (0.1 mL) was added into the mixed solution, the reaction vessel is protected from the light and under nitrogen atmosphere, mixture solution stirred for 120 minutes. After this the solvent was removed *in vacuo*, the product re-dissolved in 1.5 mL water and purified by semi preparative C12 HPLC by increasing the gradient from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes. This was followed by the continuous wash with 20% acetonitrile (0.05% TFA) for a further 20 minutes. The fractions containing product were combined and lyophilised to yield the purified peptide as a white solid. The peptide was characterised by ES-MS and analytical C12 HPLC.

44AA ESI-MS: expected: M=5091.7, *m/z* observed: $[M+6H]^{6+}$=849.5, $[M+7H]^{7+}$=728.4, $[M+8H]^{8+}$=637.4, $[M+9H]^{9+}$=566.7, $[M+10H]^{10+}$=510.1, $[M+11H]^{11+}$=463.9, consistent with a deconvoluted M=5091.4.

48AA ESI-MS: expected: M=5744.4, *m*/*z* observed: $[M+6H]^{6+}$=958.3, $[M+7H]^{7+}$=821.6, $[M+8H]^{8+}$=719.0, $[M+9H]^{9+}$=639.2, $[M+10H]^{10+}$=575.4, consistent with a deconvoluted M=5744.0.

50AA ESI-MS: expected: M=6291.2, *m*/*z* observed: $[M+5H]^{5+}$=1259.6, $[M+6H]^{6+}$=1049.6, $[M+7]^{7+}$=899.9 $[M+8H]^{8+}$=787.5, $[M+9H]^{9+}$=700.1, $[M+10H]^{10+}$=630.2, $[M+11H]^{11+}$=573.0, $[M+12H]^{12+}$=525.4, consistent with a deconvoluted M=6292.2.

52AA ESI-MS: expected: M=6405.3, *m*/*z* observed: $[M+7H]^{7+}$=915.7, $[M+8H]^{8+}$=801.5, $[M+9H]^{9+}$=712.8, $[M+10H]^{10+}$=641.6, $[M+11H]^{11+}$=583.3, $[M+12H]^{12+}$=534.7, $[M+13H]^{13+}$=493.6, consistent with a deconvoluted M=6404.7.

56AA ESI-MS: expected: M=6633.5, *m*/*z* observed: $[M+6H]^{6+}$=1106.7, $[M+7H]^{7+}$=948.88, $[M+8H]^{8+}$=830.39, $[M+9H]^{9+}$=738.12, $[M+10H]^{10+}$=664.51, consistent with a deconvoluted M=6634.7.

7.4.2 Monitor the coupling reaction of cysteine using the Ellman's test

The coupling of cysteine with 1,1'-bis(hydroxymethylene) ferrocene was monitored by the Ellman's test.[89] Aliquots of the reaction mixture were monitored prior to the addition of TFA into the mixture solution, and 17 min, 30 min, 60 min, 75 min, 90 min, 24 hour and 50 hour of the reaction time, were monitored by the Ellman's test (see section 7.2.4).

7.4.3 Monitor the coupling reaction of 28AA and 26AA by HPLC and Ellman's test

The Ellman's test of the 28AA reaction mixtures were recorded at prior to and after the addition of TFA after 10 min, 63 min, 75 min, 103 min, 215 min, 310 min and 480 min of the reaction, were monitored by the Ellman's test (see section 7.2.4).

The reaction was also monitored by analytical reversed-phase C12-HPLC by increasing the gradient from 100% water (0.05% TFA) to 20% acetonitrile (0.05 % TFA) over 20 minutes. This was followed by the continuous wash with 20% acetonitrile (0.05% TFA) for a further 20 minutes. The 26AA coupling solution was monitored by analytical HPLC before the addition of TFA, 50 min, 110 min, 160 min of reaction time, after 225 min, when a further 1/6 equivalent 1,1'-bis(hydroxymethylene) ferrocene was added, and finally after 280 minutes.

7.4.4 Extinction coefficient determination of Ferrocene

The extinction coefficient of ferrocene was determined by recording the UV-visible spectra of 1,1'-s-ferrocenyldimethyl-L-cysteine and 1,1'-s-ferrocenyldimethyl-L-glutathione at various concentrations. A stock solution of 1,1'-s-ferrocenyldimethyl-L-cysteine was diluted to 1.84×10^{-2} mM, 3.69×10^{-2} mM, 5.53×10^{-2} mM and 7.37×10^{-2} mM in water and their UV-visible spectra recorded between 200 nm and 300 nm. A stock solution of 1,1'-s-ferrocenyldimethyl-L-glutathione was diluted to 4.05×10^{-2} mM, 6.08×10^{-2} mM and 8.09×10^{-2} mM in water and their UV spectra were recorded between 200 nm and 300 nm. The experiments were carried out in triplicate to validate the data. Spectra of both complexes display a peak centered around 255 nm.

The average absorbance values at 255 nm were plot as a function of concentration, from the slope of which their extinction coefficients was determined.

7.4.5 Extinction coefficient determination of ferrocene peptide dimer

For the UV of 26AA, two 2.6 mM and one 2.4 mM in water stock solutions were prepared and quantified by the Ellman's test. Solutions of 1.7 mM, 1.3 mM and 0.97 mM 26AA were prepared by dilution of the stock solution and their UV spectra were recorded between 200 nm and 300 nm. The experiment was carried out in triplicate to validate the data.

For the UV of 28AA and 24AA, a 3.55 mM, a 3.6 mM and a 3.75 mM in water stock solution were prepared and quantified by the Ellman's test. 0.22, 0.42, 0.63 and 0.75 mM 28AA solutions were diluted from the stock solutions and their UV spectra were recorded between 200 nm and 300 nm. The experiments were carried out in triplicate to validate the data.

7.5.1 Electrochemical study of 1,1'-bis(hydroxymethylene)ferrocene and 1,1'-s-ferrocenyldimethyl-L-glutathione

The bis-substituted molecules' electrochemistry was studied by cyclic voltammetry and differential pulse voltammetry.

All the glassware and vial lids were cleaned using the following procedures: soaking in 20% H_2SO_4 aqueous solution overnight; rinsed with ultra-pure water (× 10); soaking in ammonia-hydrogen peroxide (1:1) solution overnight, rinsed with ultra-

pure water (× 10), soaking in ultra-pure water overnight, rinsed with water (× 10) and then dried in a clean oven exclusively used for electrochemistry glassware.

Before each scan of sample, the Ag/AgCl reference electrode was rinsed with water; the palladium counter electrode was burned with a Bunsen burner, the working electrode was polished by 1, 0.3 and 0.05 micron alumina powder for 8 minutes respectively followed by 10 ultra-pure water rinses and sonications, a blank buffer solution was scanned in 12 segments (6 CV scans) to confirm the experimental set up is clean.

The cyclic voltammetry (CV) was recorded on a CH Instruments Electrochemical Analyser with the Initial and low E of 0 V, High E of 0.7 V, 0.1 V/s scan rate and the sensitivity of $5e^{-7}$ A/V.

The differential pulse voltammetry (DPV) was recorded on an Autolab from -0.05 V to 0.7 V at the scan rate of 3 mV/s.

7.5.2 CV titration of 1,1'-bis(hydroxymethylene)ferrocene and ,1'-s-ferrocenyldimethyl-L-glutathione solutions

A 600 µL solution of water, 100 mM NaCl and 100 mM phosphate buffer pH 7 (60 µL) was prepared and degased by nitrogen gas for 30 minutes in a narrow-bottom conical glass vial within a stirrer bar to raise the liquid level for adequate immersion of electrodes. The experimental set up was kept in a CN Instruments Picoamp Booster & Faraday Cage. A gold working electrode was used in the experiment.

A stock solution of 1,1'-bis(hydroxymethylene)ferrocene (1.6 mg) in water (1 mL) was titrated into the buffer solution, 1 µL at a time until it gave a sufficiently strong CV signal to be analysed. After each addition the solution was degased for 30 minutes before recording a CV.

The solution of water (8 mL), 1M NaCl (1 mL) and 0.1M phosphate buffer pH 7 (1 mL) was prepared and degased by nitrogen gas for 30 minutes in a 22 mL glass vial. 1, 1'-s-Ferrocenyldimethyl-L-glutathione (1 mg) was dissolved in 0.1M phosphate buffer pH 7 (1.2 mL) to prepare 1 mM stock solution. The 1, 1'-s-ferrocenyldimethyl-L-glutathione dilutions of 1, 0.5, 0.25, 0.125, 0.0625 mM were added into the 200 µL sample holder held in the solution in a glass vial, and degased for 20 minutes before recording the CV with either a gold or carbon working electrodes.

7.5.3 DPV of 1,1'-s-ferrocenyldimethyl-L-glutathione solutions

The DPV of a 1,1'-s-ferrocenyldimethyl-L-glutathione solution were recorded twice for 2 different stock solutions with a gold working electrode, and a low volume 200 µL sample holder kept in a vial within buffer solution (15 mL, 0.1 M phosphate buffer pH 7, 0.1 M NaCl).

A stock solution (7.14 mM) was prepared by dissolving 1,1'-s-ferrocenyldimethyl-L-glutathione (1 mg) in 0.1 M phosphate buffer pH 7 (170 µL). All dilutions were degased for 30 minutes prior to recording the DPV scan. In experiment 1, the DPV was recorded for the 7140, 178, 35.7 and 17.9 µM dilutions. In experiment 2, the DPV was recorded for the 84, 47, 23, 11.5 and 5.7 µM dilutions.

7.5.4 CV and DPV of ferrocene peptide dimers by carbon and gold electrode

All the glassware was cleaned as is described in section 7.5.1. Before each scan of sample, the Ag/AgCl reference electrode was rinsed by water; the palladium counter electrode was burnt by Bunsen burner, the working electrode was polished by 1, 0.3 and 0.05 micron alumina powder for 8 minutes respectively followed by water rinse and sonication. A blank buffer solution was scanned in 12 segments to confirm the kit is clean.

The peptide solution was degased by nitrogen gas for 30 minutes before each sample scan.

In set A (SVC-3), the working electrode was placed into a sample holder which contained the ferrocene peptide dimers in 170 µL 0.01 M phosphate buffer pH 7 and 0.1 M NaCl in pure water; the sample holder, reference and counter electrodes were put into the sample vial with 10 mL 0.01 M phosphate buffer pH 7 and 0.1 M NaCl in pure water solution.

In the CV of 52AA Fc by set A, initial E is 0 V, high E is 0.7 V and low E is 0 V; 0.1 mM dimers were scanned at 5, 10, 20, 50, 100, 200 and 500 mV/s by gold and carbon electrode. The experiment was repeated in triplicate to validate the data.

The DPV titration of 52AA Fc was carried in set A kit by gold electrode. A stock solution of 52AA Fc was titrated into a 170 µL solution of 0.1 M NaCl and 0.01 M phosphate buffer pH 7 to make 12.85, 17, 22.7, 31 and 40 µM dilutions. The DPV was scanned from -0.05 V to 0.7 V at 3mV/s

In the CV of 52AA Fc by set B, the solution of water (480 µL), 1M NaCl (60 µL) and 0.1M phosphate buffer pH 7 (60 µL) was prepared and degased by nitrogen gas for 30 minutes in a narrow-bottom conical glass vial within a stirrer bar to rise the liquid level for the electrodes. The experiment kit was kept in a CN Instruments Picoamp Booster & Faraday Cage. Gold working electrode was used in the experiment. Peptide dimers (44AA Fc, 48AA Fc, 50AA Fc, 52AA Fc and 56AA Fc) were added to the vial to dilute to 100 µM for CV (initial E is 0 V, high E is 0.7 V and low E is 0 V) at 20 mV/s.

7.5.5 CV of ferrocene peptide dimers with CRE and NS DNA

All the glassware was cleaned as is described in section 7.5.1. All the CV experiments were conducted on a CH instruments Electrochemical Analyser with a three electrode system consisting of a gold working electrode, a Ag/AgCl reference electrode and a platinum counter electrode.

Before each scan of sample, the Ag/AgCl reference electrode was rinsed by water; the palladium counter electrode was burned by Bunsen burner, the working electrode was polished by 1, 0.3 and 0.05 micron alumina powder for 8 minutes respectively followed by 10 times ultra-pure water rinse and sonication, a blank buffer solution was scanned in 12 segments (6 CV scans) to confirm the experimental set up is clean.

In the CV of ferrocene peptide dimers in the presence of DNA, the solution of water (360 µL), 1M NaCl (60 µL), 500 µM double stranded DNA (120 µL) and 0.1M phosphate buffer pH 7 (60 µL) was prepared and degased by nitrogen gas for 30

minutes in a narrow-bottom conical glass vial with a stirrer bar to raise the liquid level in order to submerge the electrodes. The experiment setup was kept in a CH Instruments Picoamp Booster & Faraday Cage. Ferrocene peptide dimers (44AA, 48AA, 50AA, 52AA and 56AA) were added to the sample so that the final concentration was 100 µM for CV (initial potential is 0 V, high potential is 0.7 V and low potential is 0 V) at 20 mV/s. All the CV experiments were independently repeated in triplicate to validate the data.

7.5.6 DPV of ferrocene peptide dimers with CRE and NS DNA

All the DPV experiments were conducted on an Autolab potentiostat with a three electrode system consisting of a gold working electrode, a Ag/AgCl reference electrode and a platinum counter electrode. All the glassware was cleaned as is described in section 7.5.1.

In the DPV, water (360 µL), 1M NaCl (60 µL), 200 µM double stranded DNA (120 µL) and 0.1 M phosphate buffer pH 7 (60 µL) was prepared and degased by nitrogen gas for 30 minutes in a narrow-bottom conical glass vial with a stirrer bar to raise the liquid level in order to submerge the electrodes. Peptide dimers (44AA, 48AA, 50AA, 52AA and 56AA) were added to the sample so that the final concentration was 40 µM for DPV (initial potential is -0.05 V, end potential is 0.7 V) at 3 mV/s. The data was processed by Autolab GPES software.

7.6 Circular dichroism of peptide monomers and ferrocene peptide dimers in the absence and presence of CRE and NS DNA.

For the circular dichroism (CD) of peptide monomers and ferrocene peptide dimers, a solution of 240 µL ultra-pure water, 30 µL 0.1 M phosphate buffer pH 7 and 30 µL 1 M NaCl in pure water, was used as blank control. For the CD of peptides in the presence of DNA, a solution of 210 µL pure water, 30 µL 0.1 M phosphate buffer pH 7, 30 µL 5 µM annealed double stranded DNA and 30 µL 1 M NaCl in pure water, was used as blank control. Concentrated stock solutions of peptide monomers and ferrocene peptide dimers were added to the blank to make a final concentration of 5 µM ferrocene peptide dimers or 10 µM peptide monomers. CD spectra were recorded between 200 and 300 nm in a 1 mm path length quartz cuvette and the spectra recorded as an average of 16 scans. All the CD experiments were repeated in triplicated with different peptide stock solutions to validate the data, see section 2.7 in chapter 2.

7.7 Molecular dynamic simulations of ferrocene peptide dimer bind to CRE DNA

The molecular dynamics (MD) models were prepared in insight software based on the modified crystal structure of complex including GCN4 and CRE DNA (PDB code: 2DGC). The excess residues on the GCN4 peptide model were removed and N-terminal groups and C-terminal Cp ring tagged cysteine were put into the correspond position on the peptide. Parameters for the ferrocene-cysteine unit were imported to define the ferrocene group structure. The solvent box TIP3PBOX 15 and sodium salt were employed to the centre of peptide-DNA complex to model the water solvent by

leap software. The simulations were all performed in energy minimisation and then MD simulation. All the simulation stages were carried out using the UoB Blue Bear computer cluster using amber version 12, with the all atom force field. In the initial 0.5 ns MD simulation, the simulation was raised to 298 K from 0 K and then continues for the rest of the simulations up to 10 ns. Each complex simulation was performed in triplicate with different ig numbers to validate the results and the average model presented as the result. The 44AA, 48AA, 52AA and 56AA complex with CRE DNA were used for MD simulations.

References

1. C. R. Calladine, H. R. Drew, B. F. Luisi, A. A. Travers, *Understanding DNA: The Molecule and How it Works*, 3rd ed., Elsevier, San Diego, **2004**.
2. E. Pazos, J. Mosquera, M. E. Vazquez, J. L. Mascarenas, *Chem. Bio. Chem.*, **2011**, 12, 1958.
3. G. B. Bauer, L. F. Povirk, *Nucleic Acids Res.*, **1997**, 25, 1211.
4. J. Miller, A. D. McLachlan,A. Klug, *EMBO J.*, **1985**, 4, 1609.
5. A. Paul, S. Bhattacharya, *Curr. Sci.*, **2012**,109, 212.
6. A. K. Todd, A. Adams, J. H. Thorpe, W. A. Denny, L. P. G. Wakelin and C. J. Cardin, *J. Med. Chem.*, **1999**, 42, 536.
7. R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. Dickerson, *Nature,* **1980**, 287, 5784.
8. B.S. Reddy, S.K. Sharma, J.W. Lown, *Curr. Med. Chem.*, **2001**, 8, 475.
9. C. Zimmer, U. Wähnert, *Prog. Biophys. Molec. Biol.*, **1986**, 47, 31.
10. Y. Parajo, J. Malina, I. Meistermann, G. J. Clarkson, M. Pascu, A. Rodger, M. J. Hannon, P. Lincoln, *Dalton Trans.*, **2009**, 7, 4868.
11. S. S. Krishna, I. Majumdar, N. V. Grishin, *Nucleic Acids Res.* **2003**, 31, 532.
12. F. R. Jackson, S. Banfi, A . Guffanti, E Rossi, *Genomics*. **1997**, 41, 444.
13. J. Font, J. P. Mackay, *Methods Mol. Bio.,* **2010**, 649, 479.
14. M. Papwoth, P. Kolasinska, M. Minczuk, *Gene*, **2006**, 366, 27.
15. D. Neuhaus, Y. Nakaseko, J. W. Schwabe, A. Klug, *J. Mol. Biol.,* **1992**, 228, 637.
16. N. P. Pavletich, C. O. Pabo, *Science*, **1991**, 252, 809.
17. M. Elrod-Erickson, M. A. Rould, L. Nekludova, C. O. Pabo, *Structure*, **1996**, 4, 1171.
18. C. O. Pabo, R. T. Sauer, *Annu. Rev. Biochem.,* **1992**, 61, 1053.
19. W. A McLaughin, H. M. Berman, *J. Mol. Biol.,* **2003**, 330, 43.
20. W. Xiong, T. Li, K. Chen, K. Tang, *Nucleic Acids Res.*, **2009**, 37, 5632.
21. A. L. Anantharaman, S. Balaji, M. M. Babu, L. M. Iyer, *FEMS Microbiol. Rev.,* **2005***,* 29, 231.

22. R. A. Albright, B. W Matthews, *J. Mol. Biol*., **1998**, 280, 137.

23. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science*, **1988**, 240, 1759.

24. K. T. O'Neil, R. H. Hoess, W. F. DeGrado, *Science*, **1990**, 249, 774.

25. K. J. Lumb, P. S. Kim, *Science*, **1995**, 268, 436.

26. J. J. Hollenbeck, M. G. Oakley, *Biochem.,* **2000**, 39, 6380.

27. K. Struhl, *TIBS*, **1989**, 14, 137.

28. E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *Scinece*. **1991**, 254, 539.

29. W. Keller, P. Konig,T. J. Richmond, *J. Mol. Biol.*, **1995**, 254, 657.

30. D. N. Paolella, C. R. Palmer, A. Schepartz, *Science*, **1994**, 264, 1130.

31. T. Wang,W. L. Lau, W. F. DeGrado, F. Gai, *Biophys J.*, **2005**, 89, 4180.

32. W. T. Pu, K. Struhl, *Proc. Natl. Acad. Sci. USA,* **1991,** 88, 6901.

33. I. A.Hopel, K. Struhl, *The EMBO Journal*, **1987**, 6, 2781.

34. J. J. Hollenbeck, D. G. Gurnon, G. C. Fazio, J. J. Carlson, M. G. Oakley *Biochem.*, **2001**, 40, 13833.

35. R. V. Talanian, C. J. McKnight, P. S. Kim, *Science*, **1990**, 249, 769.

36. A. M. Caamano, M. E. Vazquez, J. M. Costas, L. Castedo, J. L. Mascarenas. *Chem. Commun.*, **2000**, 39, 3104.

37. M. Ueno, A. Murakami, K. Makino,, T. Morri, *J. Am. Chem. Soc.*, **1993**, 115, 12575.

38. B. Cuenoud, A. Schepartz, *Proc. Natl. Acad. Sci. US.*, **1993**, 90, 1154.

39. A. G. Hinnebusch, *Annu. Rev. Microbiol.,* **2005**, 59, 407.

40. R. Zenz, R. Eferl, C. Scheinecker, K. Redlich, J. Smolen, H. B. Schonthaler, L. Kenner, E. Tschachler, E. F. Wagner. *Arthritis Research & Therapy*, **2008**, 10, 201.

41. J. Liu, Q. Zheng, Y. Deng, C. S. Cheng, N. R. Kallenbach, M. Lu, *Proc. Natl. Acad. Sci. USA*, **2006**, 103, 15457.

42. E. K. O'shea, K. J. Lumb, P. S. Kim, *Curr. Biol.*, **1993**, 3, 658.

43. H. C. Hurst, *Protein Profile*, **1994**, 1**,** 135.

44. J. W. Sellers , A. C. Vincent, K. Struhl, *Mol. Cell. Biol.*, **1990**, 10, 5077.

45. A. Sitlani, D. M. Crothers, *Proc. Natl. Acad. Sci. USA,* **1996**, 95, 1404.

46. H. C. Hurst, *Protein Profile*, **1994**, 1, 140.

47. T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell*, **1992**, 71, 1223.

48. J. Kim, D. Tzamarias, T. Ellenberger, S. C. Harrison, K. Struhl, *Proc. Natl. Acad. Sci. USA,* **1993**, 90, 4513.

49. W. T. Pu, K. Struhl, *Proc. Natl. Acad. Sci. USA,* **1991,** 88, 6901.

50. A. R. Lajmi, M. E. Lovrencic, T. R. Wallace, R. R. Thomlinson, J. A. Shin, *J. Am. Chem. Soc.*, **2000**, 122, 5638.

51. G. H. Bird, A. R. Lajmi, J. A. Shin, *Biopolymers*, **2002***,* 65, 10.

52. T. Sera, P. G. Schultz, *Proc. Natl. Acad. Sci. USA*, **1996**, 93, 2920.

53. W. T. Pu, K. Struhl, *Proc. Natl. Acad. Sci. USA,* **1991**, 88, 6901.

54. J. E. Ladbury and M. L. Doyle, *Biocalorimetry 2: applications of calorimetry in the biological sciences*, Wiley, England, **2004**.

55. S. Marti´c, M. Labib, P. O. Shipman, H.B. Kraatz, *Dalton Trans*., **2011**, 40, 7264.

56. P. D. Beer, P. A. Gale, G. Z. Chen, *J. Chem. Soc., Dalton Trans.*, **1999**, 1897.

57. H. Plenio, R. Diodone, *J. Organomet. Chem.*, **1995**, 492, 73.

58. P. D. Beer, Z. Chen, A. J. Pilgrim, *J. Chem. Soc., Faraday Trans.*, **1995**, 4331.

59. P. D. Beer, Z. Chen, A. J. Pilgrim, *J. Electroanal. Chem.*, **1998**, 444, 209.

60. P. D. Beer, Z. Chen, M. G. B. Drew, J. E. Kingston, M. I. Ogden, P. Spencer, *J. Chem. Soc., Chem. Commun.*, **1993**, 1046.

61. M. E. Padilla-Tosta, R. Martínez-Máñez, T. Pardo, J. Soto, M. J. L. Tendero, *Chem. Commun.*, **1997**, 887.

62. J. M. Lloris, R. Martínez-Máñez, T. Pardo, J. Soto, M. E. Padilla-Tosta, *Chem. Commun.*, **1998**, 837.

63. H. Plenio, C. Aberle, *Chem. Commun.*, **1998**, 2697.

64. P. D. Beer, C. Blackburn, J. F. McAleer, H. Sikanyika, *Inorg. Chem.*, **1990**, 29, 378.

65. A. Gourdon, *New J. Chem.*, **1992**, 16, 953.

66. P. D. Beer, J. P. Danks, D. Hesek, J. F. McAleer, *J. Chem. Soc., Chem. Commun.*, **1993**, 1735.

67. J. C. Medina, T. T. Goodnow, M. T. Rojas, J. L. Atwood, B. C. Lynn, A. E. Kaifer G. W. Gokel, *J. Am. Chem. Soc.*, **1992**, 114, 10583.

68. P. D. Beer, M. G. B. Drew, D. K. Smith, *J. Organomet. Chem.*, **1997**, 543, 259.

69. H. Xiao, L. Liu, F. Meng, J. Huang, G. Li, *Anal. Chem.*, **2008**, 80, 5272.

70. N. Zhao, Y. He, X. Mao, Y. Sun, X. Zhang, C.-Z. Li, Y. Lin, G. Liu*, Electrochem. Commun.*, **2010**, 12, 471.

71. E. P. Diamandis, *Clin. Chem.*, **2000**, 46, 896.

72. A. Stangelberger, M. Margreiter, C. Seitz, B. Djavan, *J. Med. Hum. Genet.*, **2007**, 4, 233.

73. K. Kerman, K. A. Mahmoud, H.-B. Kraatz, *Chem. Commun.*, **2007**, 37, 3829.

74. K. Kerman, H.-B. Kraatz, *Analyst*, **2009**, 134, 2400.

75. C. C. G. Scully, P. Jensen, P. J. Rutledge, *J. Organomet. Chem.*, **2008**, 693, 2869.

76. B. Hong, R. Nauss, I. M. Harwood, S. M. Miller, *Biochemistry*, **2010**, 49, 8187.

77. P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science*, **1991**, 254, 1497.

78. A. Maurer, H. B. Kraatz, N. Metzler-Nolte, *Eur. J. Inorg. Chem.*, **2005**, 3207.

79. A. Paul, R. M. Watson, E.Wierzbinski, K. L. Davis, A. Sha, C.Achim, D. H. Waldeck, *J. Phys. Chem. B*, **2010**, 114, 14140.

80. H. Aoki, H. Tao, *Anal. Sci.*, **2008**, 24, 929.

81. D. Kwon, K. Kim, J. Kwak, *Electroanal.*, **2008**, 20, 1204.

82. X. Luo, T. M. Lee, I. Hsing, *Anal. Chem.*, **2008**, 80, 7341.

83. S. C. Hillier, C. G. Frost, A. T. A. Jenkins, H. T. Braven, R. W. Keay, S. E. Flower, J. M. Clarkson. *Bioelectrochem.*, **2004**, 63, 307.

84. R. B. Merrified, *J. Amer. Chem. Soc*, **1963**, 85, 2149.

85. A. Stadler, C. O. Kappe, *Eur. J. Org. Chem.*, **2001**, 2001, 919.

86. CEM, Microwave Synthesis: γ-Lactam Formation Studies, Application Note BIO-0006.

87. S. A. Palasek, Z. J. Cos, J. M. Collins, *J. Pept. Sci.*, **2007**, 13, 143.

88. D. L. Pavia, G. M. Lampman, G. S. Kriz. *Introduction to spectroscopy 3rd Ed*, Thomson Learning, **2001**.

89. G. L. Ellman, *Arch. Biochem. Biophys.*, **1959**, 82, 70.

90. R. W. Riddles, R. L. Blakeley, B. Zerner, *Meth. Enzymol*., **1983**, 91, 49.

91. A. Shimizu, M. Ikeguchi, S. Sugai, *J. Biomol. NMR.*, **1994**, 4, 859.

92. E. I. Solomon, A. B. P. Lever, *Inorganic electronic structure and spectroscopy*, Wiley-interscience, Toronto, **2006**.

93. G. D. Fasman, *Circular dichroism and the conformational analysis of biomolecules*, Springer, New York, **1996**.

94. K. Nakanishi, N. Berova, R. Woody, *Circular dichroism: principles and applications*, Wiley –VCH, New York, **1994**.

95. D. H. A. Correa, C. H. I. Ramos, *Afr. J. Biochem. Res.*, **2009**, 3, 164.

96. N. Greenfield, G. D. Fasman, *Biochem.*, **1969**, 8, 4108.

97. A. C. Fisher, *Electrode dynamics*, Oxford science publications, **2006**.

98. J. A. V. Butler, *Trans. Faraday Soc.* **1924**, 19, 729.

99. T. E. Gruz, M. Volmer, *Z Phys. Chem A-Chem. Kinet. Elecktrochem. Eigensch. lehre.*, **1930**, 150, 203.

100. F. G. Cottrell, *Z. Phys. Chem.*, **1903**, 42, 385.

101. F. A. Settle, Handbook of instrumental techniques for analytical chemistry, Prentice hall, USA, **1997**.

102. D.R. Staveren, N. Metzler-Nolte, *Chem. Rev.*, **2004**, 104, 5931.

103. L. Barisic, V. Rapic, N. Metzler-Nolte, *Eur. J. Chem.*, **2006**, 20, 4019.

104. K. Heize, U. Wild, M. Beckmann, Eur. *J. Inorg. Chem.*, **2007**, 4, 617.

105. S. V. Sudhir, P. NY. Kumar, S. Chandrasekara, *Tetrahedron*, **2010**, 66, 1327.

106. O. Brosch, T. Weyhermuller, N. Metzler-Nolte, *Inorg. Chem.*, **1999**, 38, 5308.

107. O. Brosch, T. Weyhermuller, N. Metzler-Nolte, *Eur. J. Inorg. Chem.*, **2000**, 2, 323.

108. K. D. Gleria, O. H. Hill, L. Wong, *Febs. Lett.*, 1996, **390**, 142.

109. K. K. Lo, J. S. Lau, N. Zhu, *Dalton Trans.*, **2002**, 8, 1753.

110. P. Haquette, M. Salmain, K. Svedlung, A. Martel, B. Rudolf, J. Zakrzewski, S. Cordier, T. Roisnel, C. Fosse , G. Jaoue, *Chem. Bio. Chem.*, **2007**, 8, 224.

111. Y. Katayama, S. Sakakihara, M. Maeda, *Anal. Sci.*, **2001**. 17, 17.

112. W. Chan, P. white. *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*. New York, **1999**, ISBN-10: 0199637245**.**

113. E. Kaiser, R. L. Colescot, C. D. Bossinge, and P. I. Cook, *Anal. Biochem.*, **1970**, 34, 595.

114. *Microwave Methods for Loading Resins*, CEM Application Note BIO-0014, 2.

115. C. Glidewell, B. J. L. Royles, D. M. Smith, *J. Org. Chem*, **1997**, 527, 259.

116. J. J. Kaminski, C. Puchalski, D. M. Solomon, R. K. Rizvi, D. J. Conn, A. J. Elliott, R. G. Lovey, H. Guzik, P. J. S. Chiu, J. F. Long, A. T. Mcphail, *J. Med. Chem.*, **1989**, 32, 1686.

117. A.Connell, P. J. Holliman, I. R. Butler, L. Male, S. J. Coles, P. N. Horton, M. B. Hursthouse, W. Clegg, L. Russo. *J. Organomet. Chem.,* **2009**, 694, 2020.

118. A. N. Patwa, R. G. Gonnade, V. A. Kumar, M. M. Bhadbhade, K. N. Ganesh. *J. Org. Chem.*, **2010**, 75, 8705.

119. R. S. Nicholson, I. *Shain, Anal. Chem*., **1964**, 36, 706.

120. S. Craz, C. Berger, A. Baici, I. Jelesarov, H. R. Bosshard. *Biochem.*, **2004**, 43, 718.

121. M. Zhang, B. Wu, H. Zhao, J. W. Taylor, *J. Pep. Sci.*, **2002**, 8, 125.

122. X. Wang, W. Cao, A. Ca, L. Lai, *Biophy. J.*, **2003,** 84, 1867.

123. K. A. Mahmoud, H. B. Kraatz, *Chem. Eur. J.*, **2007**, 13, 5885.

124. K. Plumb, H. B. Kraatz, *Bioconjugate. Chem.*, **2003**, 14, 601.

125. N. Husken, M. Gebala, W. Schuhmann, N. Metzler-Nolte, *Chem. Bio. Chem.*, **2010**, 11, 1754.

126. M. A. Weiss,, T. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison, K. Struhl, *Nature*, **1990**, 347, 575.

127. Solomons, Graham, Craig Fryhle. Organic Chemistry. 9th ed. USA: John Wiley & Sons.

128. J. G. Lee, A. J. Miles, F. Wien and B. A. Wallace, *Bioinformatics*, **2006**, 22, 1955.

129. C. Toniolo, A. Polese, F. Formaggio, M. Crisma and J. Kamphuis, *J. Am. Chem. Soc.*, **1996**, 118, 2744.

130. Z. Biron, S. Khare, A. O. Samson, Y. Hayek, F. Naider and J. Anglister, *Biochem.* **2002**, 41, 12687.

131. *European Pharmacopoeia, section* 2.2.41, **2005**, 63.

132. J. D. Morrisett, J. S. David, H. J. Pownall, A. M. Gotto, *Biochem.*, **1973**, 12, 1290.

133. J. J. Hollenbeck, D. L. Mcclain, M. G. Oakley, *Prot. Sci.*, **2002**, 11, 2740.

134. S. Oommachen, J. Ren, C. M. McCallum, *J. Phys. Chem. B*, **2008**, 112, 5702.

135. N. Yumoto, S. Murase, T. Hattori, H. Yamanoto, Y. Tatsu, S. Yoshikawa, *Biochem. Biophys. Res. Commun.*, **1993**, 196, 1490.

136. B. Blazy, F. Culard, J. C. Maurizot, *J. Mol. Biol.*, **1987**, 195, 175.

137. F. Connor, P. D. Cary, C. M. read, N. S. Preston, P. C. Driscoll, P. Denny, C. Crane-Robinson, A. Ashworth, *Nucleic Acids Res.*, **1994**, 22, 3339.

138. D. Meierhans, M. Sieber, R. K. Allemann, *Nucleic Acids Res.*, **1997**, 25, 4537.

139. P. Konig, T. J. Richmond*, J. Mol. Biol.,* **1993**, 233, 139.

140. M. Nakayama, T. Ihara, K. Maeda, *Talanta*, **2002**, 56, 857.

141. D. N. Beratan, J. N. Onuchic, J. R. Winkler, H. B. Gray, *Science*, **1992**, 258, 1740.

142. T. D. Halazonetis, K. Georgopoulos, M. E. Greenberg, P. Leder, *Cell*, **1988**, 55, 917.

143. T. K. Kerppola, T. Curran, *Cell*, **1991**, 66, 317.

144. G. Vamosi, Nina Baudendistel, C. V. D. Lieth, N. Szaloki, G. Moscar, G. Muller, P. Brazda, W. Waldeck, S. Damjanovich, J. Langowski, K. Toth, *Biophy. J.*, **2008**, 94, 2859.

145. L. Chen, J. N. Glover, P. G. Hogan, A. Rao, S. C. Harrison, *Nature*, **1998**, 392, 42.

146. T. Morii, M. Simomura, S. Morimoto, I. Saito, *J. Am, Chem, Soc.*, **1993**, 115, 1150.

147. P. C. Ma, M. A. Rould, H. Weintraub, C. O. Pabo, *Cell*, **1994**, 77, 451.

148. J. C. Forgie, S. E. Khakani, D. D. MacNeil, D. Rochefort. *Phys. Chem. Chem. Phys.*, **2013**, 15, 7713.

149. D. Lednicer, C. R. Hauser, *Org. Synth.,* **1960**, 40, 45.

150. N. H. Evans, C. H. Serpell, P. D. Beer, *Chem. Comm.*, **2011**, 47, 8775.

151. K. Heize, M. Schlenker, *Eur. J. Inorg. Chem.*, **2004**, 14, 2974.

152. S. Masaru, E. Seiji, A. Sadatoshi, *Synthesis*, **1981**, 6, 472.